Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Optimisation of Gene editing for Cystic Fibrosis

Thesis presented by

Karen Mention, MSc

for the degree of

Doctor of Philosophy

University College Cork

Department of Physiology

Head of Department: Prof. Ken O'Halloran

Supervisors: Dr. Patrick Harrison, Dr. Martina Scallan

Submitted the 6th of June 2020

TABLE OF CONTENTS

DECLARATION

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

TABLE OF FIGURES

Figures:

Tables:

ABSTRACT

Cystic Fibrosis (CF) is a recessive genetic disease caused by mutations in the *Cystic Fibrosis Transmembrane conductance Regulator* (*CFTR)* gene. To date, 352 variants in the *CFTR* gene have been shown to be CF-causing. CF is the most common genetic disease in Caucasian population, with an estimation of about 70,000 to 100,000 people living with CF worldwide. The disease results in premature death at a median age of 44 years old, with patients dying mostly from end-stage lung disease as a consequence of chronic lung infections. There is no cure for CF, but there are a range of drugs to treat CF symptoms. Over the last nine years, some small molecule drugs called modulators, were designed to improve the processing and function of the CFTR protein slowing the progression of the disease for more than 90% of CF-patients. Even though those modulators revolutionised CF treatment, the cost for those treatments are expensive, cumbersome and there are still 10% of patients with no specific drug. Indeed, some CF-causing mutations, classified as Class I variants, result in expression of little or no CFTR protein; protein modulator therapies are ineffective for patients suffering from such mutations. The variant W1282X is one of them. The W1282X variant is the 6th most common CF-causing variant, concerning 2.5% of CF patients, moreover, it is the 2nd most common class I variant. Since the discovery of the *CFTR* gene in 1989, it was expected that being able to treat the genetic problem, could lead to a treatment for CF. Since then, multiple clinical trials for *CFTR* cDNA addition have been performed, unsuccessfully. However, since the discovery of programmable nucleases, for gene editing, new hopes for CF gene therapy emerged. Indeed, some clinical trials are in process for other diseases such as Leber's congenital amaurosis, haemophilia B or mucopolysaccharidosis I and II.

The goal of this project was to compare four different techniques to correct the W1282X mutation, either by itself using homology-directed repair (HDR) and base editing, or as a superexon to correct this mutation and all the ones downstream. The purpose was to determine if there was one technique that was optimal for CF correction.

Targeting single mutations, the results showed that high correction efficiencies (around 20% with SpCas9 HDR and base editing and 8% with AsCas12a HDR) could be achieved, and the corrections led to accumulation of corrected mRNA (50% for AsCas12a HDR and Base editing to 60% for SpCas9 HDR).

In addition, CFTR protein expression could also be observed in AsCas12a-edited samples. However, using HDR, a large amount of indels could be detected, disrupting the *CFTR* gene in non-corrected alleles. Moreover, base editing showed formation of by-stander modifications within the window of editing. Using a superexon for *CFTR* correction, the homology-independent targeted integration (HITI) technique showed an intermediate level of correction efficiency of about 6% in 16HBE14o$^-$ cells after selection, leading to about 8% of corrected mRNA. Using HDR to replace a large DNA sequence, the efficiency without selection appeared to be low with about 0.02% of mRNA correction; editing at DNA level could not be determined for this technique in the cell lines available. Even though the efficiencies appeared to be lower using a superexon, the systems seemed to be safer with indels localised in introns.

Using those data, it could be possible to have a clear understanding of different gene editing techniques to correct the W1282X mutation. Those techniques could be used for other mutations as well as for other genetic diseases. With further optimisation, one or many of these techniques could be tested on CF animal models to provide safety data for a potential future use in the clinic for CF-patients.

AKNOWLEDGEMENT

I would like to thank my Supervisor Dr. Patrick Harrison, as well as my colleagues and friends Kader Cavusoglu-Doran, Lucia Santos, Elena Rojas and David Sanz for accepting me in their laboratory and helping me all along my PhD by teaching me and having productive discussions.

I would also like to thank the CF Trust for funding this PhD, which was a great experience for me.

I would like to thank CF Ireland as well, for the travel grant that gave me the opportunity to present my work at the ASGCT conference in May 2020, giving me the chance to meet many important scientists in my field of research, with potential job opportunities.

In addition, I am very grateful for the European Cystic Fibrosis Society for giving me the opportunity to present my work at the ECFS 2019 conference in Croatia.

I would like to thank my cosupervisor Dr. Martina Scallan, for taking the time to check on my well-being all along my PhD.

Thank you for my Progress Review Panel members Vincent Healy (UCC) and Therese Ruane-O'Hora, for helping me and giving me great advices all along my PhD.

I would like to thank Dr. Niall Hyland (APC, UCC) and Maximilian Woodall (PhD student in St George's University, London) for training me on the Ussing chamber experiments.

Moreover, I would like the thank Panagiota Stamou (APC, UCC) for the FACS experiments.

I am also very grateful for Dr. Robert Tarran (UNC), for taking time to review my introduction on ASL and SPLUNC1, as well as Steven Blagden (World Precision Instruments) and Pawan Jolly (Harvard) for taking their time answering my questions and helping me to understand the TEER principle.

Finally, I would like to thank all my family and my friends for the mental support, and the motivations they gave me to finish this PhD.

ABBREVIATIONS

AAV – **A**deno-**A**ssociated **V**irus

ABE – **A**denine **B**ase-**E**ditor

ASL – **A**irway **S**urface **L**iquid

BER – **B**ase **E**xcision **R**epair

Cas – **C**RISPR-**as**sociated protein

CBE – **C**ytidine **B**ase **E**ditor

CF –**C**ystic **F**ibrosis

CFTR – **C**ystic **F**ibrosis **T**ransmembrane conductance **R**egulator

CMV – **C**yto**M**egalo**V**irus

Cpf1 – **C**RISPR from ***p****revotella* and ***f****rancisella***1**

CRISPR – **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats

crRNA – crispr RNA

dCas9 – **d**eactivated Cas9

dHJs – **d**ouble **H**olliday **J**unctions

DNA-PK$_{CS}$ – **P**rotein **K**inase **c**atalytic **s**ubunit

DSB – **D**ouble-**S**tranded **B**reak

EMA – **E**uropean **M**edicines **A**gency

EMG – **E**xpression **M**ini**G**ene

ENaC – **E**pithelial sodium (**Na**$^{+}$) **C**hannel

EVOM – **E**pithelial **V**olt**O**hm**M**eter

FACS – **F**luorescence-**A**ctivated **C**ell **S**orting

FDA – **F**ood and **D**rug **A**dministration

gRNA – guide RNA

HBE cells – **H**uman **B**ronchial **E**pithelial cells

HDR – **H**omology-**D**irected **R**epair

HEK cells – **H**uman **E**mbryonic **K**idney cells

HITI – **H**omology-**I**ndependent **T**argeted **I**ntegration

HR – **H**omologous **R**ecombination

ICE – **I**nference of **C**RISPR **E**dits

Indels – **in**sertions/**del**etions

MMR – **M**is**M**atch**R**epair

MSD – **M**embrane **S**panning **D**omains

NBD – **N**ucleotide **B**inding **D**omain

nCas9 – **n**ickase Cas9

NGS – **N**ext **G**eneration **S**equencing

NHEJ – **N**on-**H**omologous **E**nd **J**oining

NMD – **N**onsense-**M**ediated **D**ecay

PAM – **P**rotospacer **A**djacent **M**otif

PCL – **P**eri**C**iliary **L**iquid

PI domain – **P**AM-**I**nteracting domain

pre-crRNA – precursor crisprRNA

PTC – **P**remature **T**ermination **C**odon

PTO – Phosphorothioate

RNP – **R**ibo**N**ucleo**P**rotein

RT – **R**everse-**T**ranscription

SaCas9 – ***S****taphylococcus* ***a****ureus* Cas9

SDBR – **D**ouble-**S**trand **B**reak **R**epair

SDSA – **S**ynthesis-**D**ependent **S**trand **A**nnealing

SE / SEv1 / SEv2 – **S**uper**E**xon / **S**uper**E**xon **v**ersion **1** /**S**uper**E**xon **v**ersion **2**

SeV – **Se**ndai **V**irus

sgRNA – **s**ingle **g**uide RNA

SPLUNC1 – **S**hort-**P**alate **L**ung and **N**asal **E**pithelial **C**lone **1**

SRC – **S**trategic **R**esearch **C**entre

ssODN – **s**ingle-**s**tranded **O**ligo**D**eoxyribo**N**ucleotides (ssODN)

TALEN – **T**ranscription **A**ctivator-**L**ike **E**ffector **N**ucleases

TEER – **T**rans-**E**pithelial **E**lectrical **R**esistance

TIDE – **T**racking of **I**ndels by **De**composition

tracrRNA – **tr**ans-**a**ctivating crisprRNA

WT – **W**ild-**T**ype

ZFN – **Z**inc-**F**inger **N**uclease

CHAPTER 1 - INTRODUCTION

1.1 CYSTIC FIBROSIS: A GENETIC DISEASE

Cystic **F**ibrosis (CF) is a recessive genetic disease, meaning that, to be sick, the person has to get faulty genes from both of their parents. Indeed, each human gets two copies of each gene, one from each of their parents. Therefore, if both parents carry a non-functional gene, ¼ babies might have CF (Figure 1.1).

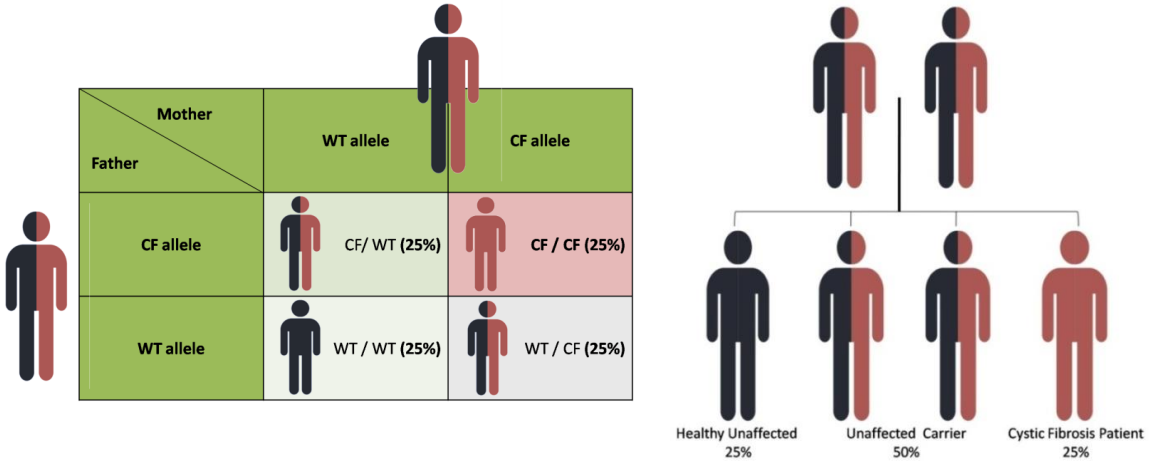


Figure 1.1: Graphic representation of the Mendelian recessive mode of inheritance of Cystic Fibrosis. Both parents have one healthy gene and one faulty gene. Following the table showing the gene distribution, there will be 25% chance for each child to get Cystic Fibrosis and 50% chance for each child to be a healthy carrier.

Figure 1.1 - Mendelian recessive mode of inheritance of Cystic Fibrosis

Cystic Fibrosis is the most common genetic disease in Caucasian populations. The last numbers given by the Cystic Fibrosis Trust (CF Trust) in the UK (1), and the **N**ational **I**nstitute of **H**ealth (NIH) in the US (2), showed an incidence of 1 baby born with CF in every 2500 or 3500 birth respectively. However, the incidence in African and Asian populations is estimated to be around 1 in 17000 and 1 in 31000 respectively (2). Overall, it is estimated that approximately 70 000 to 100 000 people live with CF worldwide (3). In Ireland, 1/19 people is a CF gene carrier, while it is estimated to be 1/25 in the UK. Ireland is the country having the highest CF incidence with 0.03% person having CF. Moreover, 1/1353 Irish baby is born with CF making Ireland far ahead from other countries with the next higher incidence countries being UK and Belgium with about 0.01% (4).

Cystic Fibrosis is also a fatal disease killing two people every week in the UK (1). Indeed, already in the 18th and 19th century, a legend was circulating that if a child's skin tasted salty, this child was cursed and was going to die. However, since the first description of CF by Dorothy Andersen in 1938 (5), the median survival improved greatly thanks to the improvement in disease understanding and treatments (Figure 1.2; (6)). However, in 1966, still 80% of the CF babies died before reaching 5 years old (7). Around 1987, the median survival increased to 8 years old and by 1994 the median survival increased again to 21 years old, making CF an adult disease (8). The latest data from the CF foundation, described a median survival of 44 years old for babies who are born between 2014 and 2018 (9).

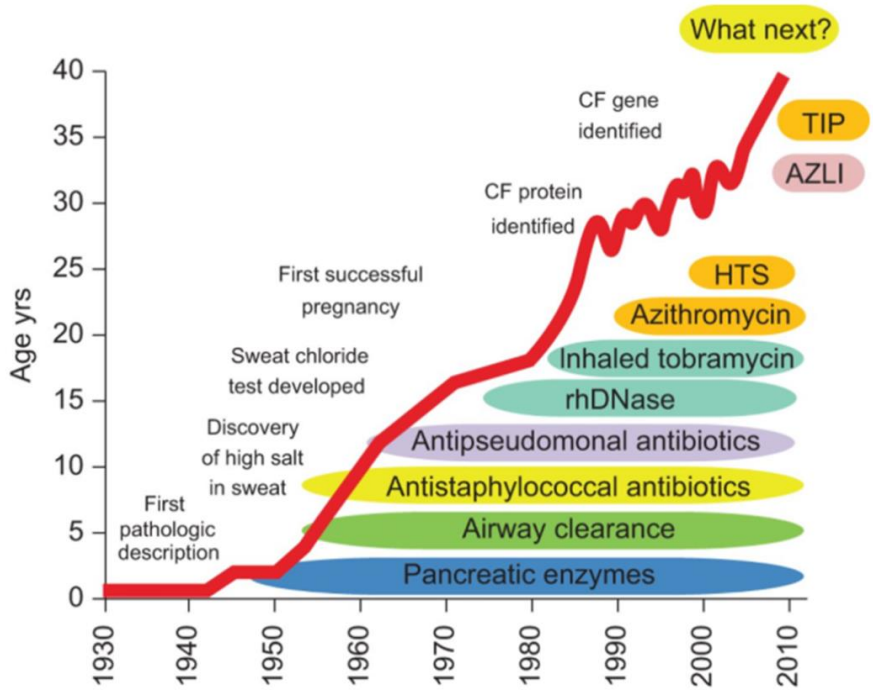


Figure 1.2: Survival age curve, for Cystic Fibrosis, since 1930. Developments in the understanding and treatments of the disease allowed a clear improvement in the survival of CF patients. ***(Figure from Elborn 2013;*** *(6)****).***

Figure 1.2 - Survival age curve, for Cystic Fibrosis, since 1930

Cystic Fibrosis is a multi-organ progressive disease, meaning that the disease affects many organs and symptoms get worse over time. Amongst the most common symptoms, pulmonary diseases such as lung obstruction and lung infection, are the most common and lethal, affecting almost 100% of the patients (10, 11).

Other symptoms include pancreas insufficiency where the duodenal content of CF patients is low in volume, thick and sticky, obstructing the pancreatic ducts where enzymes get stuck and activate prematurely. All this lead to inflammation, fibrosis and permanent scarring of the pancreas. This deficiency in enzymes' secretion, lead to CF being unable to digest nutrients properly (12, 13). Older patients also develop diabetes mellitus (14). Moreover, other deadly symptoms affect the liver and are called **CF L**iver related **D**iseases (CFLD). Those diseases include obstructive jaundice and cirrhosis (10, 15). Amongst all of this, CF also affects the intestines and reproductive organs with reports of bowel obstruction in about 35% of patients (10), and male infertility (16). Patients with CF display a high concentration of salt in their sweat (average of 106 mEq./l) compared to healthy controls (average of 32 mEq./l) (17). This feature is often used as a diagnostic for the disease (17, 18). CF disease is due to mutations in the *CFTR* gene, however, some other factors, genetic or environmental, can dictate the severity of the symptoms (Figure 1.3).

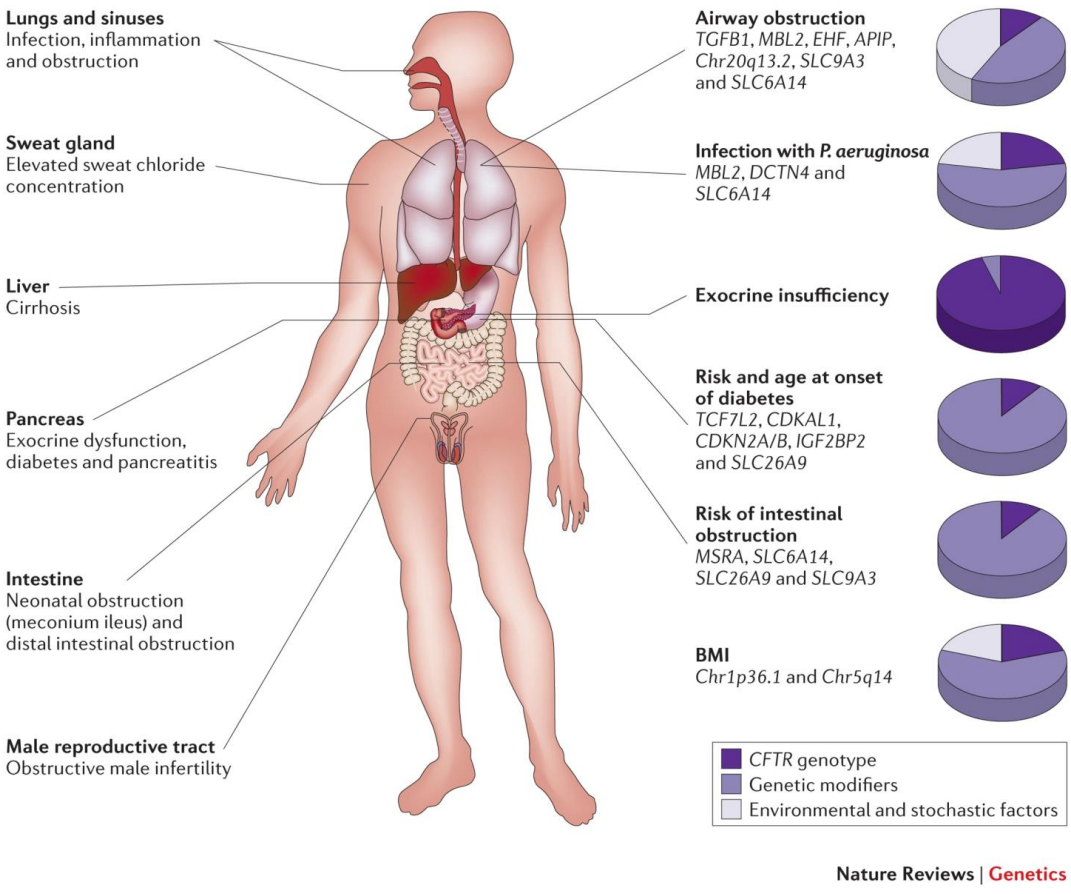


Figure 1.3: CF symptoms and contribution to genetics and environmental factors. ***(Figure from Cutting 2015;*** *(19)****)***.

Figure 1.3 - CF symptoms and contribution to genetics and environmental factors

1.2 DEFECT OF THE CFTR CHANNEL IN EPITHELIA

It is important to identify the defects in cystic fibrosis, in order to understand the origin of the symptoms. In 1989, the ***C****ystic* ***F****ibrosis* ***T****ransmembrane conductance* ***R****egulator* (*CFTR*) gene was discovered (20, 21). Disease-causing mutations in the gene are responsible for a decrease or loss of function of the CFTR protein, leading to the CF symptoms described above. In the next sections, an overview of the protein synthesis and structure will be detailed, followed by the protein involvement in the disease and how it explains the CF symptoms.

1.2.1 - CFTR PROCESSING: FROM GENE TO PROTEIN

The *CFTR* gene is first transcribed into a messenger RNA (mRNA) inside the nucleus. Interestingly, early on after the discovery of the *CFTR* gene, it was noticed that the exon 10, (originally exon 9; Figure 1.6), was often spliced out of the *CFTR* mRNA. After studies of the splicing sites around the exon, it was discovered that an adenine was deleted from the splicing donor sequence in intron 10. Indeed, instead of the consensus splice donor sequence 5'-GTAAGTT-3', the splice donor sequence in the *CFTR* gene was 5'-GTAGTT-3'. This modification from the consensus sequence leads to a misplicing that results in exon 10 loss. The resulting protein from this misplicing is misfolded and non-functional (22).

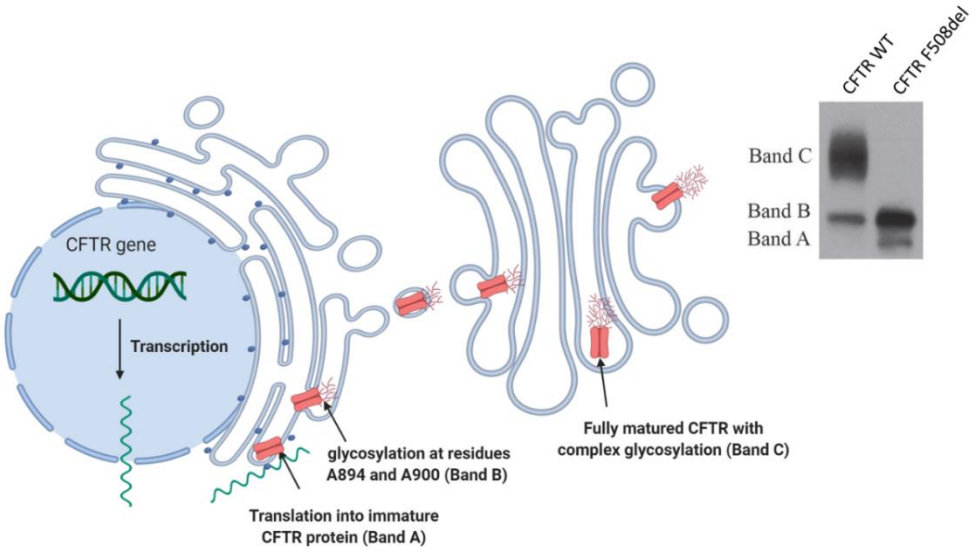
After transcription, the *CFTR* mRNA undergoes translation, initially as immature CFTR protein in the endoplasmic reticulum. This immature protein can be detected by western blot and is a low molecular weight band known as Band A (130kDa; Figure 1.4A). The newly synthesised protein undergoes a first glycosylation process inside the endoplasmic reticulum. It has been shown that the residues Asn894 and Asn900 in the 4th extracellular loop are the ones being glycosylated (Figure 1.4B; (23, 24)). This glycosylation process increases the weight of the protein, so it migrates slower on the western blot gel, called Band B (150kDa; Figure 1.4A). The correct folding and glycosylation will allow a quality/control system inside the ER to confirm its further migration and process into the Golgi. Finally, to have the fully matured CFTR, the glycosylated protein migrates to the Golgi apparatus where it will undergo complex glycosylations. This fully matured CFTR protein is shown with an even higher molecular weight in the western blot gel and is called Band C (170kDa; Figure 1.4A). The smears observed for band C are believed to be different degrees of glycosylation.

The fully matured protein is then sent to the apical membrane of epithelial cells, where it will function as a channel for anions and bicarbonate (Figure 1.4A; (23-26)).

1.2.2 - CFTR PROTEIN STRUCTURE

The CFTR protein is anchored inside the apical membrane of epithelial cells, through the **M**embrane **S**panning **D**omains (MSD). CFTR possesses two MSDs, each of them is composed of six transmembrane domains. Each of those MSDs is linked to a **N**ucleotide **B**inding **D**omain (NBD) (Figure 1.4B). Each of those MSD-NBD motifs is separated by a unique regulatory domain and both NBDs and the regulatory domain are located intracellularly (Figure 1.4B). The MSD**s** play a role in the specificity of the CFTR pore for anions and bicarbonate. The hydrolysis of ATP by the NBDs is responsible for the gating of the channel. When there is no ATP, CFTR is opened in the intracellular side, to allow the cellular anions and bicarbonate to enter inside the pore. Once ATP binds to NBDs, there is a dimerisation of the two NBDs, which induce a conformation change to open the pore and free the molecules into extracellular side. The hydrolysis of ATP will then bring the conformation back to its original state. Finally, phosphorylation of the regulatory domain by c**A**MP-dependent **P**rotein **K**inase (PKA), controls and regulates the channel activity (25, 26), (27, 28).

A)



B)

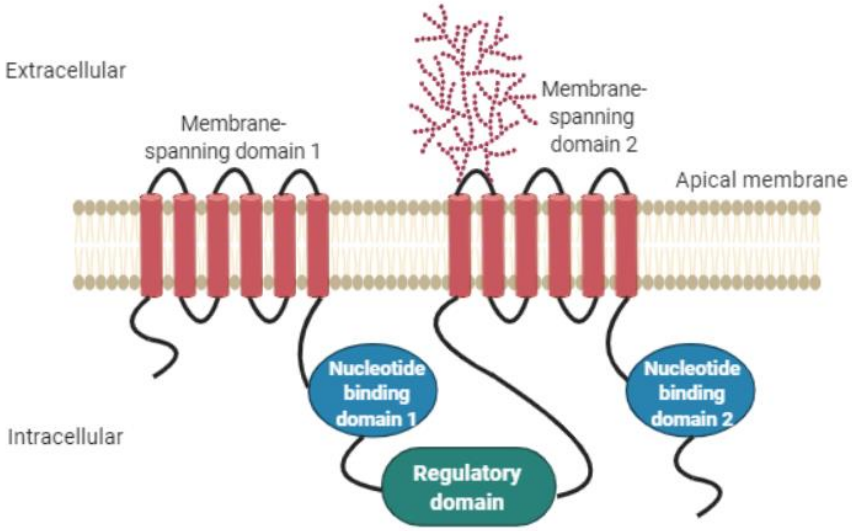


Figure 1.4: CFTR protein's processing and structure. A) CFTR protein synthesis and process through the endomembrane system. After transcription and translation, a first glycosylation process happens in the endoplasmic reticulum. The glycosylated protein migrates into the Golgi apparatus where it will undergo a more complex glycosylation. In the top right corner is shown a western blot, displaying the CFTR protein in a WT cell or a cell with the CFTR F508del mutation, which prevents CFTR processing (23). *Three different bands can be seen, representing the degree of CFTR glycosylation from non-glycosylated (Band A), first step of glycosylation (Band B) and complete glycosylation (Band C). B) Structure of the CFTR protein. The CFTR protein is an epithelial channel composed of two membrane-spanning domains (MSD), each of them composed of six transmembrane domains. Each MSD is linked to a nucleotide binding domain (NBD) localised intracellularly. The both MSD-NBDs are separated by a regulatory (R) domain localised intracellularly.*

Figure 1.4 - CFTR protein synthesis, processing and structure

1.2.2a - Physiological role of CFTR

Since the 80s, it was suspected that the cause for CF was a generalised epithelial defect. Indeed, it was shown that CF symptoms seemed to be due to an electrolyte transport dysfunction, characterised by an hyperabsorption of salt and water leading to dehydrated and sticky fluids that led to fibrosis or infections (13, 29). The problem in electrolytes and water content was observed in the secretions of diverse organs all over the organism (29).

Epithelial cells are present throughout the whole body where they create a barrier between the outside and the inside. They are present in many organs containing tunnels, ducts or cavities. Those include organs such as lungs and airways, pancreas, intestine and liver. Therefore, the link between the symptoms and the CFTR localisation can be observed. The CFTR channel is responsible for secretion of both chloride and bicarbonate ions into the lumen. Luminal fluidity is important for physiological processes such as digestion or mucociliary clearance (30). Recently, newly identified cells in the epithelium airways, called ionocytes, were shown to express more than 50% of the total lung *CFTR* mRNA. Those cells compose a small subset of pseudostratified cells representing about 1% of airway cells (31, 32).

1.2.2b - In healthy airways

Airways infections are the leading cause of death in CF. Focusing on the airway physiology, CFTR protein is responsible for the secretion of anions and bicarbonate into the **A**irway **S**urface **L**iquid (ASL). The ASL is a thin watery layer localised at the surface of airway epithelium between epithelial cells and the air in the lumen. It is composed of two layers: 1) a **P**eri**c**iliary Liquid (PCL) layer of about 7µm with a low viscosity composition. The PCL layer is localised around the cilia and separate the epithelia from the mucus. 2) A mucus layer with a gel-like composition of about 7 to 70µm. This layer contains mucins and is more viscous than the first layer. Its gel-like structure traps exogenous inhaled particles such as microorganisms and debris. The ASL is very important for defence against infections and contains many components such as antibacterial peptides, migratory innate immune cells (neutrophils, macrophages, etc…) or signalling molecules (such as cytokines). The PCL layer surrounds the cilia, which beat to remove the mucus out of the lungs, in a mechanism called

mucociliary clearance. This mechanism is vital to keep the lungs as a sterile environment (30, 33)

Consequently, a regulation of the volume and composition of this ASL is critical to keep the organism healthy. CFTR and the **E**pithelial sodium ($\mathbf{Na^+}$) **C**hannel (ENaC) are the main regulators of the ASL volume. Indeed, CFTR leads to an increase in ASL volume through chloride secretion, and ENaC leads to a decrease in ASL volume through sodium absorption (33). The secretion of chloride and bicarbonate by CFTR was shown, over the years, to have many important roles in human physiology. Indeed, bicarbonate was shown to keep the ASL at a neutral pH. Keeping the ASL's pH neutral was proved to have many important physiological functions such as keeping proteins in the ASL functional (those proteins being pH-dependent) or modifying the mucin conformation to make it loose and slippery (34). Some proteins in the ASL have very important roles to prevent lung infections, such as lysozyme, lactoferrin, defensin (34) and SPLUNC1 protein, which are antimicrobial proteins. SPLUNC1 was also demonstrated to be an inhibitor of ENaC, keeping the ASL well hydrated by preventing ENaC from absorbing too much sodium (35, 36). SPLUNC1, by its ENaC regulation, allows cilia to remove the mucus-containing bacteria from the lungs.

Chloride secretion is very important to keep a thin and watery mucus, by allowing a passive water transport through the epithelium. Indeed, the secretion of chloride into the ASL allows a flow of sodium and water through the paracellular pathway, following osmotic driving forces (Figure 1.5A; (30, 37)).

The regulation of CFTR is determined by the presence of chloride in the cell. The chloride absorption is dictated by the presence of a NKCC1 cotransporter, at the basal side, which transport two molecules of chloride, along with a sodium and potassium cation. The sodium is secreted back into the interstitial fluids by a Na^+/K^+-ATPase pump, which absorbs a potassium cation in exchange. This potassium is being secreted back into the interstitial fluids through a potassium channel (Figure 1.5A). The secretion of sodium and potassium out of the cells creates a negative gradient of chloride which will be secreted out of the cells, into the lumen, through CFTR.

1.2.2c - In CF airways

In the case of CF, the CFTR protein has reduced to no function. Because of this, the composition and volume of luminal fluids are impaired, leading to pathophysiology of many organs such as intestinal obstruction, pancreas insufficiency or lung infection. The lack of CFTR function prevents chloride and bicarbonate secretion in luminal fluids. Therefore, they become more acidic and viscous (30, 34).

In CF airways, the increased in viscosity is due to two main parameters: 1) the lack of chloride secretion does not create an osmotic flow of water, which dehydrates the ASL. 2) The ASL becomes acidic due to the lack of bicarbonate secretion that does not buffer the protons anymore. Because of this, SPLUNC1, which is a pH-dependent inhibitor of ENaC, is not functional anymore, making ENaC hyperfunctional (35). The hyperactivation of ENaC increases sodium absorption, creating a reverse passive flow of water from the ASL towards the basal side of the epithelium (38). Moreover, the increase of sodium in the cells, overactivate the Na^+/K^+-ATPase pump transport (39) (Figure 1.5B).

All these phenomena combined decrease the volume of the ASL and increase the pH and viscosity of the mucus. The sticky mucus traps bacteria and blocks the cilia from beating the mucus out of the lungs. Therefore, bacteria are stuck in this sticky mucus, antimicrobial agents are not functional because of the increase in pH, and all of this create a perfect environment for pathogens such as *Staphylococcus aureus, Pseudomonas aeruginosa* or other opportunistic pathogens to develop and create chronic lung infections that are difficult to treat (Figure 1.5B).

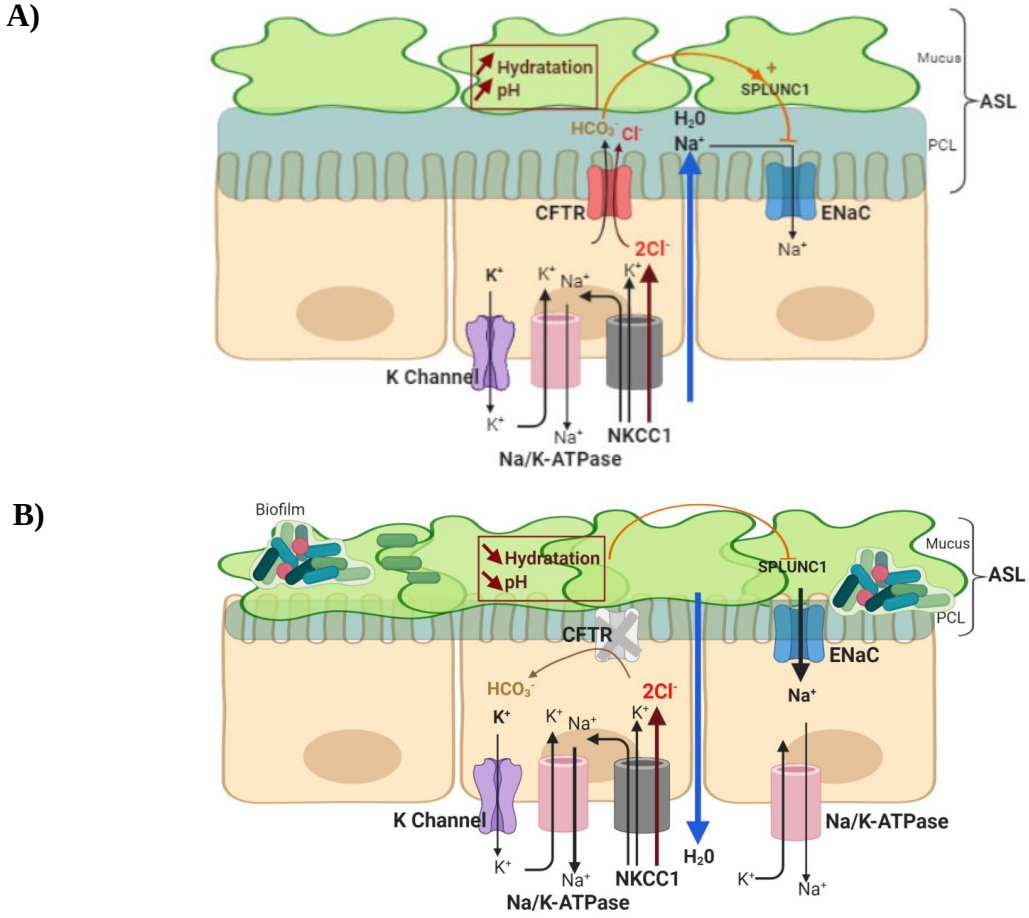


Figure 1.5: A) ***Electrophysiology of healthy airway epithelium.*** *The NKCC1 cotransporter allows entrance of one molecule of sodium and potassium with two molecules of chloride in the cells. This cotransporter is the main source of chloride into the cells. While the Na/K-ATPases and the K channels transport the sodium and potassium outside of the cells through the basal side, there will be creation of a chloride gradient that will lead to chloride secretion into the lumen by the CFTR channel. Moreover, CO_2 and H_2O inside the cells, undergo a chemical reaction that will produce bicarbonate $HCO3^-$. These $HCO3^-$ molecules will be also secreted by CFTR. The accumulation of chloride in the PCL leads to a passive flow of water that will hydrate the ASL. Bicarbonate in the ASL buffers protons to increase the pH. The increased pH activates SPLUNC1 which inhibits ENaC sodium absorption.*
B) ***Electrophysiology of CF airway epithelia.*** *In CF, CFTR channel activity is reduced or non-functional, preventing secretion of both bicarbonate and chloride in the lumen. Thus, the ASL becomes dehydrated and more acidic. This inhibits the water movements from the interstitial fluids into the lumen of the airways. Moreover, the decrease in pH inactivates SPLUNC1, which, consequently, does not inhibit ENaC. Thus, there is an increase in sodium absorption, which creates an osmotic gradient, leading to a flow of water from the ASL to the interstitial fluids. The dehydrated ASL is responsible for an increase in mucus viscosity and stickiness. The mucus blocks the cilia beats, preventing them from clearing the mucus out of the lungs. The steady sticky mucus traps bacteria which develop into biofilms and create chronic lung infections difficult to treat.*

Figure 1.5 - Electrophysiology of healthy and CF airway epithelia

1.3 THE *CFTR* GENE

1.3.1 - GENE NOMENCLATURE

The *CFTR* gene is localised on chromosome 7 in the q31 region of the long arm (20, 21). Originally, twenty four exons were characterised but after years of studies it was clear that the *CFTR* gene contains 27 exons and encodes an anions and bicarbonate-specific channel of 1480 amino acids (40, 41). The old and new exons' nomenclature is shown in Figure 1.6.

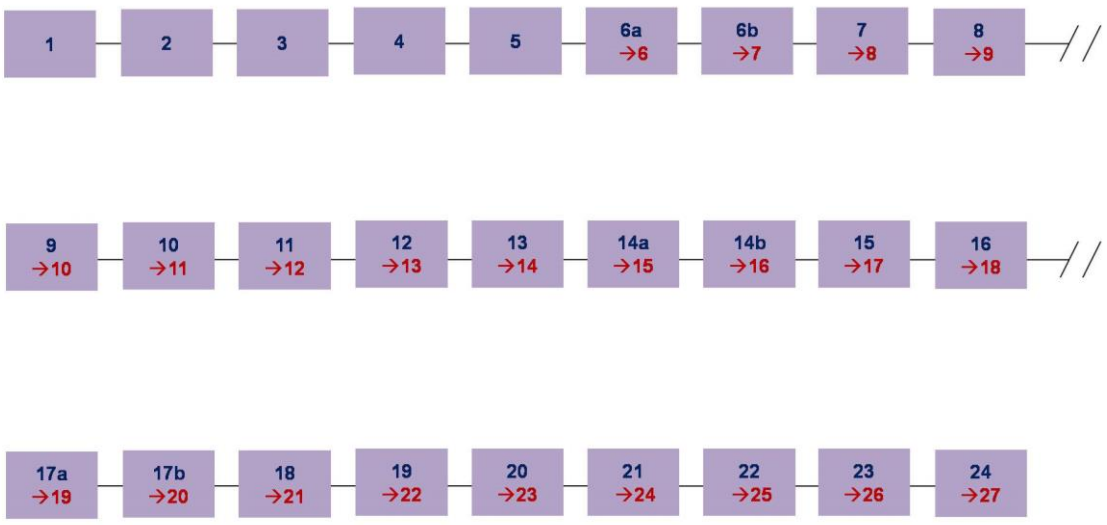


Figure 1.6: Schematic representation of the CFTR exons with the old nomenclature in blue and the new nomenclature in red.

Figure 1.6 - Schematic representation of the *CFTR* exons

Mutations in the *CFTR* gene lead to the expression of a non-functional protein responsible for the CF symptoms. Over the years, more than 2000 variants localised all over the *CFTR* gene have been discovered.

1.3.2 - *CFTR* VARIANTS NOMENCLATURE

CF variants have two nomenclatures. The old nomenclature was originated from the Cystic Fibrosis Mutation Database (CFTR1), initially created by the Cystic Fibrosis Genetic Analysis Consortium in 1989 (42), and labelled the positions of the variants starting from the beginning of exon 1.

Later on, around 2010, the nomenclature guideline changed to follow the recommendations proposed by the Human Genome Variation Society (43). The guideline recommended to use the A from the 5'-ATG-3' start codon, as the nucleotide number one.
Using the new nomenclature, if the variant concerns the cDNA sequence, the variant's name will start with a "c". The same way, if the variant concerns the protein sequence, the name will start with a "p". For example, the famous F508del (or ΔF508) mutation is now labelled c.1521_1523delCTT, at the cDNA level, or p.Phe508del, at the protein level. Similarly, the mutation W1282X is now labelled c.3846G>A, at the cDNA level, or p.Trp1282X, at the protein level.

Most research groups used the new nomenclature for exons but the old nomenclature for mutations, so that format is used in this thesis.

1.3.3 - *CFTR* VARIANTS AND CLASSIFICATIONS

In 2020, three hundred fifty-two of these variants have been shown to cause Cystic Fibrosis (44). The most common mutation, affecting about 70% of the CF patients, is a deletion of 3 base pairs in the exon 11 of the gene, leading to the deletion of a phenylalanine at the position 508 of the protein (F508del) (20). The next most common mutations are G542X (around 2.5% of the population have at least one allele with this mutation), G551D (~2.1%), N1303K (~1.5%), R117H (~1.3%) and W1282X (~1.2%) (44). The other mutations affect less than 1% of the population and some mutations are even very rare in the CF population (affecting just few patients worldwide). Those latter mutations are therefore very hard to study and not prioritised for drug development. Amongst the 2000 variants, the major deleterious ones are: 1) missense mutations (39%), 2) frameshift (16%), 3) splicing mutations (11%) and 4) nonsense mutations (8%). The less common variants account for in-frame insertions/deletions, large insertions/deletions and mutations of the promoter. Finally, the remaining of variants correspond to sequence variations, which have not yet been confirmed as disease-causing or not (42).

To facilitate the development of drugs, some groups divided the variants into different classes according to their functional defect.

This strategy aimed to better understand how variants cause disease, which may, in turn, help in designing a therapy that would target a group of mutations, compared to single mutation treatment that would necessitate more than 300 different treatments, one for each mutation (45-47).

The first classification in 2005 by Eitan Kerem, grouped the CF-causing mutations into five classes, as followed (Figure 1.7):

Class I mutations are nonsense mutations that create a premature stop codon. In that case, the mRNA is sent for degradation prematurely by **N**onsense-**M**ediated **D**ecay (NMD), resulting in very few to no CFTR proteins at the cell surface. Class I mutations account for about 7% of the population (48).

Class II mutations impair the correct folding of CFTR proteins, which remain stuck in the endoplasmic reticulum and are subsequently degraded. The result is very few to no proteins at the cell surface. Class II mutations include the most famous mutation F508del, and they account for at least 85% of the population (48).

Class III mutations impair the regulation of the CFTR protein, resulting in presence of proteins at the cell surface but with a reduced activity. In this class of mutations, the gating of the CFTR channel is very low with the protein being rarely open. Class III mutations account for about 3% of CF-patients (48).

In **class IV mutations**, the conductance of CFTR is reduced. The protein is present at the cell surface, but a problem in phosphorylation decreases the amount of chloride and bicarbonate going through. Class IV mutations account for about 3% of CF-patients (48).

Class V mutations correspond to splicing defects and account for about 3% of CF-patients (48). Modifications of splicing sites can lead to new exons, which can add amino acids and can also create premature stop codons. In this group of mutations there can be both normal and abnormal splicing and both mRNAs will be present. The normal mRNA will lead to a normal functional protein at the cell surface, while the aberrant transcript can lead to either an abnormal, non-functional protein or an unstable mRNA that will be sent for degradation through the NMD. Certain mutations can lead to a more or less strong splicing site, meaning that some mutations will lead to more or less aberrant transcripts varying the severity of the symptoms. Also, the splicing efficiency can be different from patients and from organs (47).

In 2013, **class VI mutations** were described, containing rarer mutations. This class includes mutations that impair the stability of CFTR for example by increasing the endocytosis of CFTR. CFTR is present at the cell membrane but because it is not stable, there will be less of them, resulting in a reduced overall activity (45).

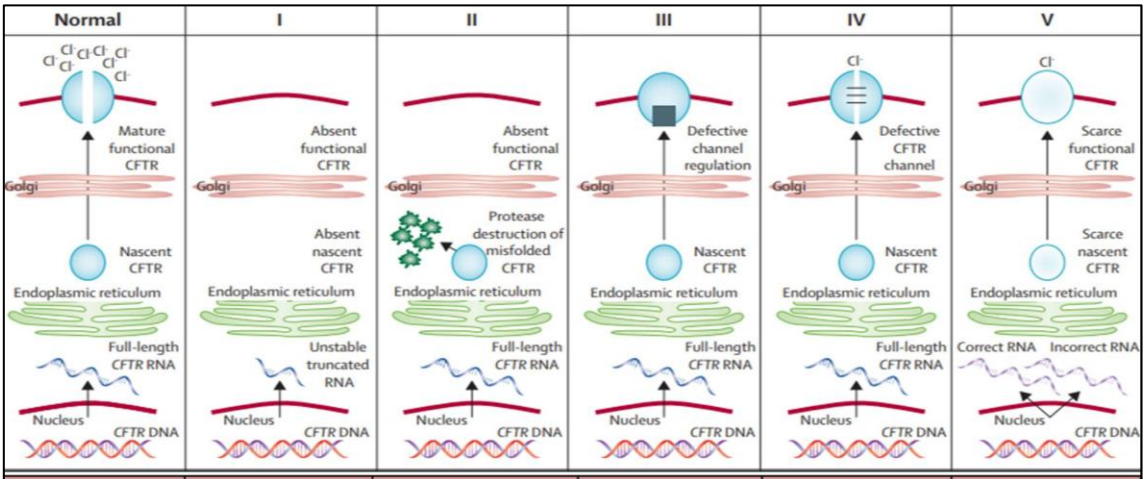
In 2016, De Boeck and Amaral described the **class VII mutations**. The logic behind this class was that the class I described mutations leading to non-stable RNA with no protein formation. However, they suggested that this class could include different kind of functional defects that would need different kind of therapies, such as premature stop codon mutations or large deletions. Therefore, in class VII, they would include every non-druggable mutation, which can be corrected only using genetics. This class included large deletions. However, this review was not in complete agreement, especially with some groups thinking that it was making the classification more confusing (49).

Each class was then termed "theratype", meaning that the disease-causing variants are classified according to the molecular-based treatment to which they are expected to respond. In regard to each theratype, potential theoretical treatments were considered and are shown in figure 1.7, along with the actual treatments available (19). To correct the CFTR protein, drugs called CFTR modulators showed a lot of interest. They are drugs that correct malfunctioning CFTR. They are composed of: 1) **correctors**, which help the correct folding and trafficking of the protein, 2) **potentiators**, which potentiate the activity of CFTR present at the cell membrane, 3) **stabilisers**, which help stabilising CFTR to increase its half-life at the apical membrane, and 4) **amplifiers**, which increase the number of CFTR protein synthesised by the cells.

For class II mutations, where the CFTR protein is not folded properly, correctors would theoretically be useful to help proteins to form properly and increase their trafficking to the cell membrane. As shown in figure 1.7, potentiators could theoretically be useful for every class of mutations except the class I. Indeed, if the defect is a low conductance or a regulation impairment, which both decrease CFTR function, potentiators could potentially increase the function of the proteins present at the membrane. Moreover, for class II mutations, potentiators alone might not have a significant function, however, once the corrector increases the amount of CFTR at the cell surface, increasing their function could theoretically make the treatment more efficient. For class I mutations, it is believed that read-through molecules which would impair the fidelity of the ribosome and jump the premature stop

codon, could be useful. However, for the time being, no such molecules have been shown to have a significant beneficial effect (50).

To finish, it is important to note that because CF is a genetic defect, every mutation could theoretically be corrected using genetics, such as gene therapy and precise genome editing. However, for the time being no gene therapy trials were successful for CF (51-54).



	Class I	Class II	Class III	Class IV	Class V
Impaired mechanism	Unstable RNA	No trafficking	Impaired regulation	Reduced conductance	Alternative splicing
Example mutations	G542X W1282X R1162X	F508del N1303K	G551D G1349D	R117H R334W	3849+10kb C->T D565G
Theoretical treatments	- Read-through molecules - Genetics	- Correctors - Potentiators - Genetics	- Potentiators - Genetics	-Potentiators - Genetics	- Potentiators - Genetics
Available treatments	None	F508del : Elexacaftor, Tezacaftor, Ivacaftor, Lumacaftor	Ivacaftor	R117H : Ivacaftor	Ivacaftor and Tezacaftor

Figure 1.7: Figure showing the five classes of mutations described by Eitan Kerem in 2005 (47). For each class, the mechanism impaired, examples of mutations, theoretical and current available treatments are listed. ***The figure has been modified from the review article from Vallieres and Elborn 2014 (55).***

Figure 1.7 - Classification of CF mutations

1.4 CURRENT TREATMENTS

Unfortunately, there is no cure for Cystic Fibrosis. However, some treatments exist to help patients having a better and prolonged life with the disease.

1.4.1 - TREATMENTS FOR THE CF SYMPTOMS

First, there are treatments to treat CF symptoms. Those treatments do not treat the CFTR defect, but help patients dealing better with the disease (Figure 1.2). For example, CF patients with pancreatic insufficiency will be prescribed some pancrelipase enzyme products, to help them digesting food and have a close to normal diet. Before those enzymes were available in the market, CF children were dying due to problems in absorbing nutrients from the food.

Because the lung phenotype is the most dangerous one, responsible for the death of most CF-patients, many treatments are prescribed to minimise the lung defect. Pulmozyme® is a treatment containing DNaseI, which hydrolyse DNA in mucus to make it thinner and easier to remove from the lungs. Moreover, hypertonic saline is also used to increase the hydration of the ASL and improve mucus clearance. To prevent chronic infections in the lungs, many antibiotics are used depending on the microbe prevalence in the airways. For example tobramycin was shown to be very efficient against *P. aeruginosa* (56).

Most of CF-patients have chronic inflammation because of their immune system constantly trying to fight bacterial infections. However, too much inflammation leads to lung damage. Therefore, research is going on to try finding some drugs that will decrease the inflammatory response, without stopping it. For the moment, only ibuprofen is used, however, two new drugs are in Phase II clinical trial, with one (LAU-7b) that finished in July 2020 and the second one (Lenabasum), which is expected to finish at the end of August 2020 (56).

1.4.2 - CFTR MODULATORS

On top of those treatments, new treatments, called CFTR modulators (described in section 1.3.3), have been described to correct the CFTR defect and are now available, with more in clinical trials (56). Those drugs are very efficient and greatly improve the lifespan of many patients. However, it targets only few mutations (about 40 mutations amongst the 352 described (57)).

The **F**ood and **D**rug **A**dministration (FDA)-approved drugs available in the market include **Trikafta™**, which is a tritherapy with a combination of two CFTR correctors (Elexacaftor and Tezacaftor) and one potentiator (Ivacaftor). This tri-therapy is efficient for patients with at least one copy of the F508del mutation, corresponding to the majority of CF-patients (90%) (56). **Kalydeco®** treatment consists of only Ivacaftor potentiator, and has been shown to be effective for 38 CF variants (58). Interestingly, amongst those variants, five are Class V mutations, one is a Class IV and all the others are class III, showing that the use of potentiators can be efficient for several classes of mutations (Figure 1.8A; (58)). **Orkambi®** treatment consists on a dual therapy of a corrector (Lumacaftor) and a potentiator (Ivacaftor). This therapy is available only for patients homozygous for F508del, which includes about 45% of CF-patients. Finally, **Symdeko®** is a dual therapy including the CFTR corrector Tezacaftor and the CFTR potentiator Ivacaftor. This treatment has been shown to correct 26 mutations including the five Class V mutations shown for Kalydeco®, and twenty one class III mutations also common to Kalydeco® (Figure 1.8B; (59)). Symdeko® treatment is prescribed for every patient that carries at least one of those 26 mutations, no matter what the other mutation is (56).

A)

Table 3: List of *CFTR* Gene Mutations that Produce CFTR Protein and are Responsive to KALYDECO					
E56K	*G178R*	*S549R*	*S977F*	*F1074L*	*2789+5G→A*
P67L	*E193K*	*G551D*	*F1052V*	*D1152H*	*3272-26A→G*
R74W	*L206W*	*G551S*	*K1060T*	*G1244E*	*3849+10kbC→T*
D110E	*R347H*	*D579G*	*A1067T*	*S1251N*	
D110H	*R352Q*	*711+3A→G*	*G1069R*	*S1255P*	
R117C	*A455E*	*E831X*	*R1070Q*	*D1270N*	
R117H	*S549N*	*S945L*	*R1070W*	*G1349D*	

Class III
Class IV
Class V

B)

F508del/F508del
c.1521_1523delCTT

OR

At least one responsive mutation from the following list:

711+3A→G c.579+3A>G	A455E c.1364C>A	D579G c.1736A>G	E193K c.577G>A	K1060T c.3179A>C	R117C c.349C>T	S945L c.2834C>T
2789+5G→A c.2657+5G>A	A1067T c.3199G>A	D1152H c.3454G>C	E831X c.2491G>T	L206W c.617T>G	R347H c.1040G>A	S977F c.2930C>T
3272-26A→G c.3140-26A>G	D110E c.330C>A	D1270N c.3808G>A	F1052V c.3154T>G	P67L c.200C>T	R352Q c.1055G>A	
3849+10kbC→T c.3718-2477C>T	D110H c.328G>C	E56K c.166G>A	F1074L c.3222T>A	R74W c.220C>T	R1070W c.3208C>T	

Class III
Class V

Figure 1.8: A) Table displaying the mutations responsive to Kalydeco®. Also displayed are the different classes in which those mutations belong. ***The table has been adapted from the Vertex website for Kalydeco®*** *(60). B) Table displaying the mutations responsive to Symdeko®. Also displayed are the different classes in which those mutations belong.* ***The table has been adapted from the Vertex website for Symdeko*** *(59).*

Figure 1.8 - Mutations responsive to Kalydeco and Symdeko treatments

Overall, it can be shown that many treatments are available for CF patients, to improve their day to day health and lifespan. However, those treatments are very expensive. For example, the new treatment Trikafta™, was estimated to cost US$311,000 per year in the USA (61). Another example in UK is the price of Orkambi being estimated to cost £105,000 per patient per year (62).

However, even though the prices for CF drugs are very expensive, they do not account for the full price of a CF patient healthcare. Indeed, a paper from 2012 estimated the annual cost for CF patients (before modulators were available) to be on average €48,603. This cost comprised hospitalisation care, medications and any other direct or non-direct healthcare.

The sum of the treatments and care altogether can lead to a significant cost burden for CF patients and relatives (63). In addition, those treatments are cumbersome to deal with, with some patients taking more than 40 tablets a day for the most severe symptoms of the disease (1). Even though CFTR modulators were shown to help most of CF-patients (more than 90%), there is still a minority of CF-patients who do not have any specific drug.

Therefore, it is important to search for treatments that will be less expensive, potentially less cumbersome and which could benefit every CF-patient, comprising those with rare mutations.

1.5 AUGMENTATION GENE THERAPY

As stated above, CF is a disease where different levels of defects lead to cumulative lung damage and death. Figure 1.9 shows the different steps of defects related to CF lung disease, as well as the different treatments that can be used for the different phases. It also shows the stages where different treatments work for the downstream symptoms, and the more upstream stages where modulators correct the protein (Figure 1.9). Acting on CFTR protein has been shown to lead to an improvement of the downstream defects (64). The success of correcting the protein with modulators has reinvigorated investigation into the possibility that maybe if the most upstream defect is corrected, which is a *CFTR* gene modification, all the following defects would be restored.

The interest in gene therapy started shortly after the discovery of the *CFTR* gene in 1989. Another reason for gene therapy interest is the potentiality to target every patient independently of their mutations, by transducing a wild-type *CFTR* complementary DNA (cDNA) into airway epithelial cells. The expectation is that if gene therapy is successful, it would lead to an amelioration of the disease in severely ill patients or prevent the development of lung disease in mildly ill patients.

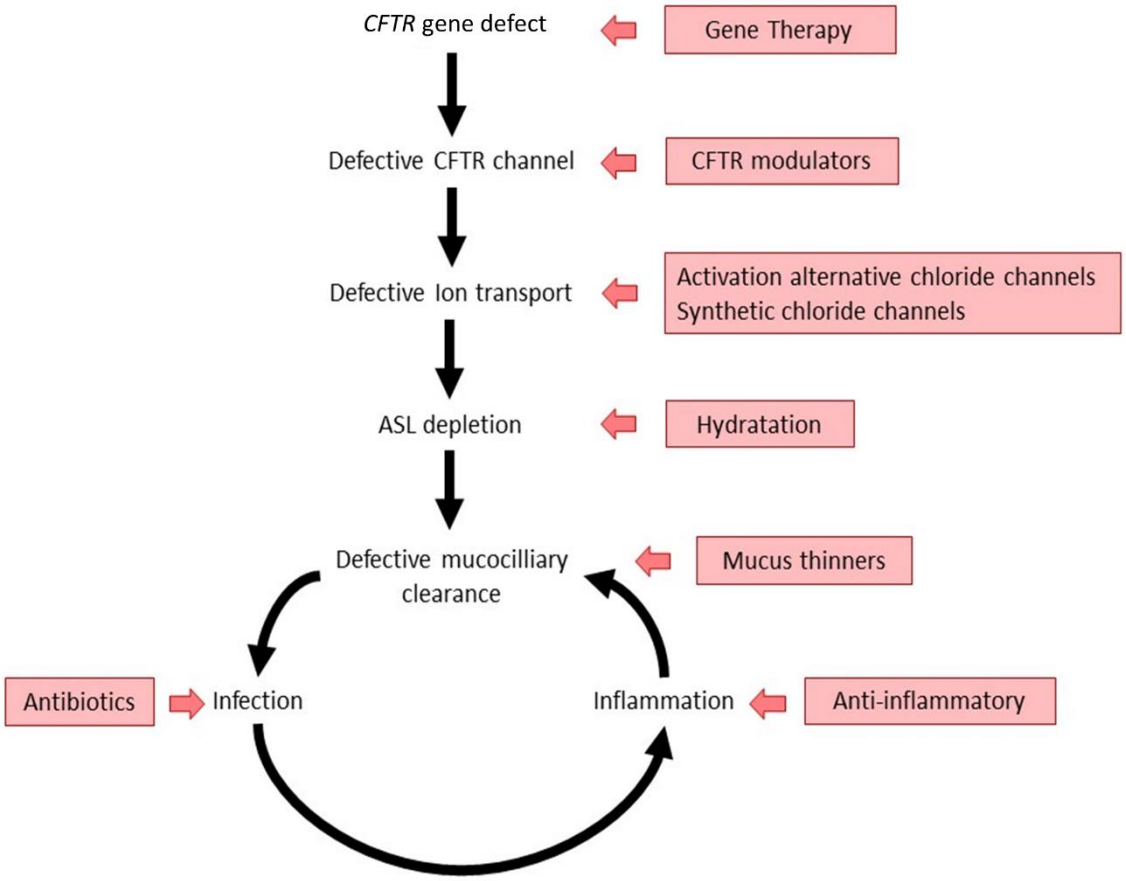


Figure 1.9: Graph representing physiological defects in CF. leading to the lung disease and the treatments associated. The defects are shown from the most upstream defect, which is gene mutations, to the most downstream defects which are the symptomatic ones.

Figure 1.9 - Physiological defects in CF and treatments associated

The **A**merican **S**ociety of **G**ene and **C**ell **T**herapy (ASGCT) defines Gene Therapy as : "the introduction, removal or change in genetic material—specifically DNA or RNA—into the cells of a patient to treat a specific disease." (65).

1.5.1 - FAILURES FOR CYSTIC FIBROSIS

Since the discovery of the *CFTR* gene in 1989, gene therapy has attracted numerous researchers with the first clinical trials starting only 4 years after the gene discovery (54). Lots of interest emerged in the 90's until early 2000's, with around 30 clinical trials in total until now. Around 28 clinical trials happened between 1993 and 2004 (see review (66)).

At that time, the gene therapy techniques used were called "augmentation gene therapy", which consisted of introducing a **W**ild-**T**ype (WT) *CFTR* cDNA in the airway epithelium

using viral (adenoviruses, lentiviruses and adeno-associated viruses (AAV)) or non-viral (cationic liposomes) vectors.

Unfortunately, the results were disappointing, and a clear loss of interest emerged after 2004. Indeed, none of the clinical trials succeeded to show a significant improvement in lung function (66-68).

Non-viral vectors were shown to be safe to administer and re-administer, however, the efficiency of transduction of airway epithelial cells was low (69). The last clinical trial with non-viral vectors was a phase 2b trial using nebulised gene-liposome complex. The results of the trial were published in 2015 and showed promising results with a modest stabilisation of the symptoms, in the treated patients, compared to a decline in the placebo group. However, the treatment effect was not sufficient to improve the lung disease and the trial did not go to the next phase (69).

Viral vectors were shown to be more efficient in transducing airway epithelial cells, however, an adaptive immunity pre-existent or forming against the viral vectors made it impossible to administer repeat doses (52, 53, 70). Indeed, neutralising antibodies targeting the viral vectors led to a decrease of vector transduction during the repeat doses. The last CF clinical trial using viral vectors was a phase 2b trial published in 2007. The trial consisted of AAV vectors encoding the *CFTR* cDNA, aerosolised into CF patients. One hundred and two CF patients were treated with either repeated doses of the viral vector, or the placebo. The results showed that repeated doses of AAV were safe. However, no improvement in lung functions were observed.

1.5.2 - GENE THERAPY TREATMENTS FOR OTHER DISEASES

Recent successes in gene therapy trials for other diseases proved that gene therapy is feasible and can be used to treat diseases. There are at least six gene therapy treatments approved by the U.S. **F**ood and **D**rug **A**dministration (FDA) and/or the **E**uropean **M**edicines **A**gency (EMA). Those treatments are Glybera®, IMLYGIC®, Strimvelis®, KYMRIAH®, YESCARTA® and LUXTURNA®.

IMLYGIC® is a medicine used to treat patients with melanoma. It consists of a vector derived from **H**erpes **S**implex **V**irus **1** (HSV1). The virus was modified to infect and kill melanoma cells only, without killing healthy cells.

The vector also encodes the **G**ranulocyte-**M**acrophage **C**olony-**S**timulating **F**actor (GM-CSF), which allows the immune system to recognise and destroy melanoma cells (71).

Strimvelis® is a treatment for patients with Severe Combined Immunodeficiency due to Adenosine Deaminase deficiency (ADA-SCID). The $CD34^+$ cells from the patient's bone marrow are genetically modified to correct the *ADA* gene, using a γ-retrovirus. The cells are infused back to the patient where the cells will produce normal lymphocytes expressing ADA (72).

KYMRIAH® and YESCARTA® are treatments use for patients with **D**iffuse **L**arge **B**-**C**ell **L**ymphoma (DLBCL). In addition, KYMRIAH® is also used for B-cell **A**cute **L**ymphoblastic **L**eukaemia (ALL) and YESCARTA® for **P**rimary **M**ediastinal large **B**-**C**ell **L**ymphoma (PMBCL). Both treatments are e*x vivo* genetic modifications of patient's white blood cells, to make them express a chimeric antigen receptor (CAR) that targets the CD19 antigen present in B-cells and absent in other cell lineages (73, 74).

LUXTURNA® is a treatment for patients with loss of vision due to biallelic RPE65 mutation, causing inherited retinal dystrophy (e.g. retinis pigmentosa and Leber's congenital amaurosis). The treatment consists of an AAV vector containing the *RPE65* gene, injected into the subretinal space to transduce retinal cells and make them express the protein (75).

Finally, **Glybera®** was a treatment used to treat patients diagnosed with familial **L**ipo**P**rotein **L**ipase **D**eficiency (LPLD). The treatment consisted of an AAV1 vector carrying the gene encoding the human Lipoprotein Lipase, under the control of a cytomegalovirus (CMV) promoter. The vector was delivered into the muscles to allow muscle cells to produce the enzyme (76).

Those gene therapy treatments used in the clinic give great hope for the development of new gene therapy techniques to treat diseases. Therefore, it gives a new optimism to develop gene therapy treatments for Cystic Fibrosis.

In 2017, however, Glybera® was withdrawn due to financial reasons. Indeed, LPLD patients are very rare, and therefore, the company uniQure withdrew the renewal of this treatment in order to focus the money on research for more common diseases such as Huntington's disease or Haemophilia B (77).

I think it is important to highlight this problem, to show the reality regarding rare diseases. Indeed, when analysing Cystic Fibrosis, there are hundreds of different disease-causing mutations, with some being common mutations (>1%) and some other being very rare (<0.001%). The patients having the latter might be left out by companies, which might want to focus their money on the most common mutations. There is therefore a need to find a treatment for CF which can help all the mutations with one treatment, to not leave any patient behind.

1.5.3 - LIMITATIONS OF AUGMENTATION GENE THERAPY

Even though there is a new optimism toward the use of augmentation gene therapy for monogenic diseases, some adverse effects observed during some earlier trials, raised some concerns in physicians and patients. Amongst those adverse effects, insertional mutagenesis is one of the most common. Between 2010 and 2012, twelve patients from four different trials, using γ-retrovirus as a gene therapy vector, developed insertional oncogenesis (78). Insertional oncogenesis is the development of malignancies after the random integration of a viral vector inside specific locations in the human genome.

1.6 GENOME EDITING

1.6.1 - PROGRAMMABLE NUCLEASES FOR PRECISE GENOME EDITING

In 1996, the group of Chandrasegaran developed a hybrid restriction enzyme using the type II restriction enzyme FokI. After studying the FokI enzyme, they discovered that it has two parts. One was a DNA binding domain, and the other was a non-specific nuclease domain (79). After this discovery, the group tried to create a hybrid by fusing other DNA binding domains, such as Zinc Finger motifs, to the FokI nuclease domain. By using Zinc Finger motifs, they hypothesised that because those motifs could be synthesised to recognise a specific sequence, a hybrid of DNA binding Zinc Finger motifs and FokI nuclease domain would be able to cut at a specific desired location (80). In collaboration with the group of Dana Carroll, the **Z**inc-**F**inger **N**ucleases (ZFN) were better characterised and used as programmable nucleases to target and cut a specific region in the genome, improving localised genetic alterations (81).

This discovery was the first step to the new era of programmable nucleases which would later comprise three groups: 1) **Z**inc-**F**inger **N**ucleases (ZFNs), 2) **T**ranscription **A**ctivator-**L**ike **E**ffector **N**ucleases (TALENs) and 3) the **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (CRISPR) / **C**RISPR-**as**sociated (Cas) proteins (CRISPR/Cas).

Programmable nucleases are powerful gene therapy tools for two main reasons: 1) they allow localised genetic alteration and improve efficiency of localised gene targeting, and 2) they decrease random integration events which are potentially dangerous in the clinic.

Those nucleases were described by the committee for advanced therapies of the European Medicines Agency as being safer than previously described gene therapies using viral vector-induced gene insertion.

They cited: "Gene correction/ gene replacement strategies aimed at targeting specific genomic sites (i.e., exploiting sequence-specific endonucleases such as artificial zinc finger or Transcription activator-like effector nuclease (TALEN)) could reduce the risk derived from random or semirandom insertion" (78).

Those programmable nucleases lead to a new gene therapy technology called "**genome editing**", described by the NIH as: "a group of technologies that give scientists the ability to change an organism's DNA. These technologies allow genetic material to be added, removed, or altered at particular locations in the genome" (82).

1.6.2 - CLINICAL TRIALS FOR PROGRAMMABLE NUCLEASES

Few clinical trials using programmable nucleases are in process. ZFNs are being used for Haemophilia B (NCT02695160), mucopolysaccharidosis I (NCT02702115) and mucopolysaccharidosis II, and CRISPR/Cas systems are being used for blindness (NCT03872479), Human Immunodeficiency Virus-infected subjects (NCT03164135) and lung cancer (NCT02793856). Even though those trials are mostly Phase I clinical trials, the process of using programmable nucleases in clinical trials gives great hopes for the scientific community that the use of genome editing can be translated to clinical use. The final completion of those trials will be between 2020 and 2022, when there will be more information about the clinical results of programmable nucleases.

1.6.3 - THREE CLASSES OF PROGRAMMABLE NUCLEASES

Jin-Soo Kim's group in 2014 published an early review comparing the three classes of programmable nucleases (83). At that time, ZFNs seemed to be the less efficient nuclease, with high rate of off-target effects and evidence of cytoxicity. The target sequences of ZFNs are largely limited to G-rich sequences and the nucleases are challenging to design. TALENs, however, have no site restriction allowing to target almost any DNA sequence. They are also very efficient at creating DSB.

However, TALENs' design can be very challenging and time consuming. Moreover, there has been evidence of recombination of the TALEN pairs within the cells. The major drawback of those two nucleases is that for every new target, a new nuclease must be designed, which is laborious and costly.

Finally, CRISPR systems have the very high advantage that their design and preparation are very simple. Indeed, because the sequence specificity is guided by an RNA molecule, the protein is always the same (see section 1.7). Therefore, to transfect cells with proteins, there is no need to purify different proteins for each target, making this system less time consuming and much cheaper to synthesise. The guide RNA (gRNA) can be designed to target a sequence of choice, making the CRISPR system easy to reprogram to target new DNA sequences (see section 1.7). However, the main inconvenience is their DNA target limitation (see section 1.7). Over the last seven years, many new versions of CRISPR have been improved to decrease this targeting limitation and more information will be given in section 1.7.

Taking all those criteria into consideration, the CRISPR systems are the one being used in this project for genetic alteration of the *CFTR* gene in CF.

1.7 THE CRISPR/CAS SYSTEMS

The discovery of CRISPR started in 1987 when a Japanese group studied the *iap* gene, responsible for alkaline Phosphatase Isozyme conversion, in *E. coli*. Analyses of the gene and its flanking sequences revealed an unusual structure at the 3' flanking end of the *iap* gene. Five homologous sequences of 29 nucleotides arranged in direct repeats were flanking 32 nucleotides unique sequences (84).

Those sequences, later called "spacers" were compared with sequences from the NCBI database and showed to be homologous to some extrachromosomal DNA from phages and plasmids (85). In 2007, the CRISPR locus was shown to confer resistance to bacteria and archaea against phages containing the exact same sequence as the spacer sequence (86). Therefore, the CRISPR locus was described as the immune system from bacteria and archaea against invading DNA, in a similar manner as RNA interference in eukaryotes (87). Around 90% of archaeal genomes and 50% of bacterial genomes were shown to possess at least one CRISPR locus (88).

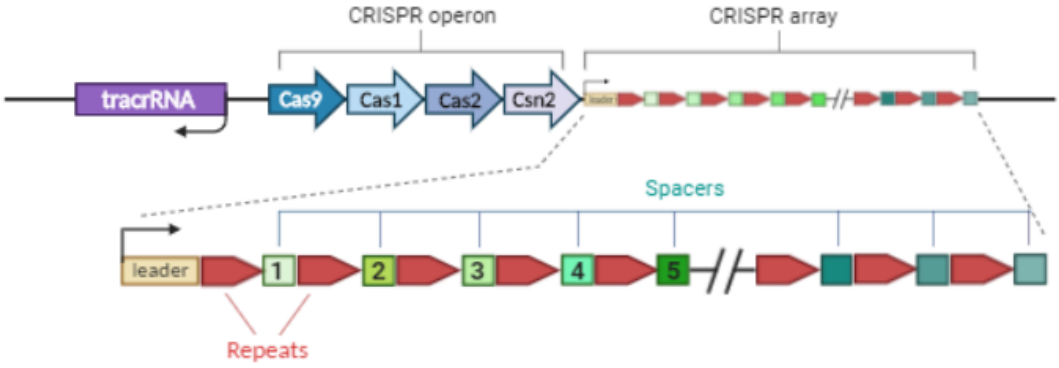
A CRISPR locus is composed of an operon of *cas* gene, followed by a CRISPR array (Figure 1.10A). The CRISPR array is composed of a leader sequence, followed by an array composed of a variable number of different spacer sequences, belonging to foreign invading DNA, interspaced with identical direct repeats of about the same size (20-50bp). The leader sequence has been shown to act as both a transcription promoter for the CRISPR array, and a signalling sequence for the insertion of new spacer sequences upon foreign DNA infection (example of a CRISPR locus for the CRISPR/Cas9 system in figure 1.10A; (89)).

This CRISPR/Cas based immune system is composed of three phases: 1) **the adaptive phase**. During this phase, foreign DNA invading the bacterium will be identified, processed into short DNA fragments (called spacers), and will be integrated into the host genome at the upstream side of the CRISPR array. 2) During **the expression phase**, the array is transcribed into a precursor CRISPR RNA (pre-crRNA). Enzymatic digestion cleaves the repeat sequences in the pre-crRNA, leading to several different short crRNAs composed of spacer sequence in 5' and a long segment from the repeat in 3'. 3) During **the interference phase**, those short crRNAs form a ribonucleic complex with Cas proteins, leading them to a complementary invading DNA (called protospacer). Together, the complex recognises and cut the protospacer, leading to a silencing of the invading DNA.

The most recent classification from 2020, described the CRISPR/Cas systems in two classes, 6 types and 33 subtypes (90). The Class I contains type I, III and IV CRISPR/Cas systems, and is characterised by the formation of a multi-subunit complex for the detection and degradation of the invading DNA (effector module, shown in orange in the figure 1.10B).

The Class II is composed of type II, V and VI CRISPR/Cas systems, and is characterised by the presence of a single effector protein responsible for recognition and degradation of the target DNA (effector module, shown in orange in the figure 1.10B).

A)



B)

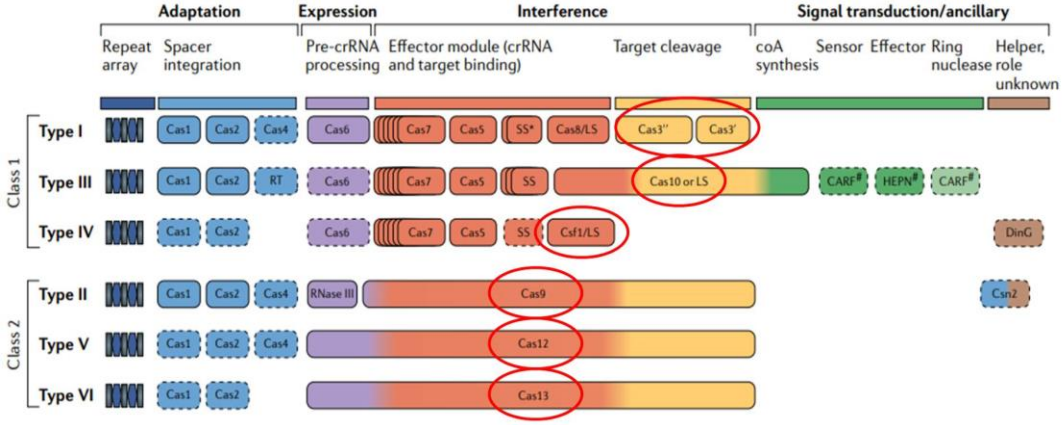


Figure 1.10: A) Graphic representation of the CRISPR locus in the case of the Class 2, Type II-A CRISPR/Cas9 system. The CRISPR locus is composed of a CRISPR operon containing several Cas genes that will have different functions in the CRISPR immune adaptation process. Downstream of the operon is the CRISPR array composed of a leader sequence, which has a promoter role, and a series of different spacers flanked by direct repeats (shown in red). Each spacer is a sequence derived from a previous DNA encounter. B) Graphic representation of the Cas effectors and their roles in the different phases of CRISPR immune adaptation, for each class and type. Noticeably, each type of CRISPR system is defined by its own characteristic "signature proteins", shown in the circles. Concerning class 2 systems, one unique protein is used for the interference phase of the CRISPR immunity. ***(Figure modified from Makarova et al. 2020;*** *(90))*.

Figure 1.10 - CRISPR locus for class 2, type II-A CRISPR/Cas9 systems, and graphic representation of the Cas effectors with their roles in the different phases of the CRISPR immune adaptation.

1.8 TYPE II CRISPR SYSTEMS

In 2012, the CRISPR/Cas systems were described for the first time as tools to create DSB for potential eukaryotic genome targeting and editing (91). For those purposes, class 2 systems were the ones the scientific community focused on, due to the simplicity of having only one single protein targeting and cleaving the sequence of interest. The best characterised class 2 system is the type II CRISPR/Cas, which is characterised by the protein Cas9. Consequently, this system is more generally called CRISPR/Cas9 system (91). CRISPR/Cas9 system is very easy to design and it has been intensively studied. Many variants of the Cas9 proteins have been created and are available to overcome sequence limitations or increase editing efficiency and specificity of targeting.

1.8.1 - CRISPR/CAS9 SYSTEM

CRISPR/Cas9 system consists of a unique effector protein (Cas9) coupled with a guide RNA (gRNA) therefore forming a complex called **R**ibo**N**ucleo**P**rotein (RNP).

1.8.1a - The Cas9 protein

Cas9 is a bi-lobed protein with one lobe being a **REC**ognition lobe (REC) and the other one being a **NUC**lease lobe (NUC). At the interface between those two lobes, there is a positively charged cleft that allows the binding between the target DNA and the gRNA (92). The REC lobe enables the recognition of the gRNA while the NUC lobe is in charge of the recognition and cleavage of the target DNA (or protospacer). The NUC lobe is composed of two nuclease domains: the HNH domain, which cleaves the target strand, and the RuvC domain, which cleaves the non-target strand of the protospacer. Moreover, the NUC lobe also contains the **P**AM-**I**nteracting (PI) domain, which recognises the **P**rotospacer **A**djacent **M**otif (PAM) on the protospacer (92).

In the case of CRISPR/Cas9, the **P**rotospacer **A**djacent **M**otif (PAM) is a short sequence localised downstream of the 20nt target DNA, on the non-target strand (Figure 1.11A). The PAM is specific to each organism. Concerning the Cas9 of the bacteria *Streptococcus pyogenes* (SpCas9), which is the most widely used for genome editing, the PAM sequence consists of three nucleotides 5'-NGG-3', "N" being any nucleotide.

1.8.1b - guide RNA (gRNA)

The gRNA composing the ribonucleoprotein, is physiologically composed of two components, the crisprRNA (crRNA) and the trans-activating crisprRNA (tracrRNA). The 42nt crRNA contains the 20nt spacer sequence followed by 22nt repeat sequence from the CRISPR array. The 89nt tracrRNA is composed, at its 3' end, of a complementary sequence to the 22nt repeat sequence of the crRNA, and a segment with a specific secondary structure localised at its 5' end (Figure 1.11B upper panel). The analysis of the dual-RNA secondary structure brought to the design of a single guide RNA (sgRNA), keeping the important structures, and replacing the crRNA:tracrRNA binding, by a synthetic linker (Figure 1.11B lower panel; (91)). The sgRNA was designed to simplify genome editing by allowing the expression of a single transcript over two distinct ones. Moreover, in the case of cell transfection with RNPs, the use of a sgRNA saves the process of having to hybridise both crRNA and tracrRNA.

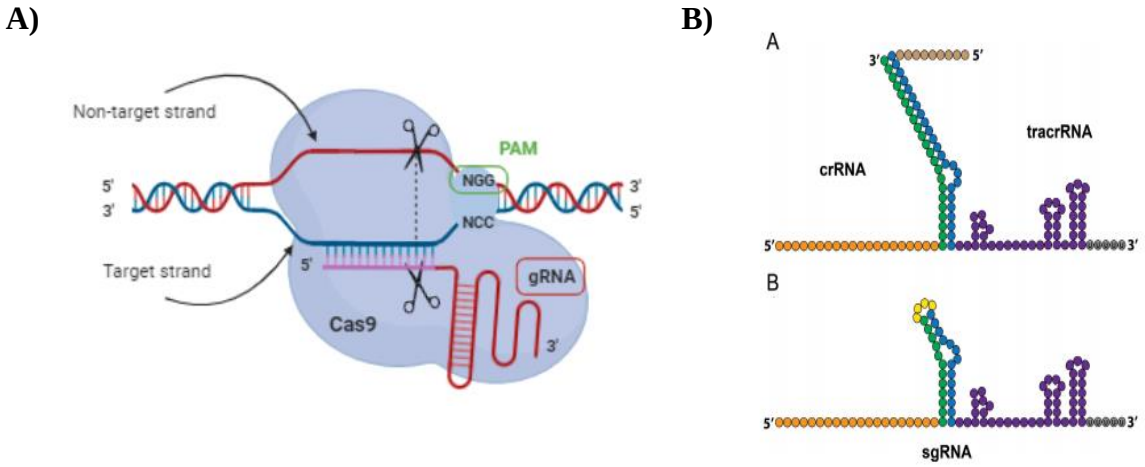


Figure 1.11: A) Graphic representation of the Cas9:gRNA ribonucleoprotein, recognising and cleaving the target DNA. The RNP recognises the PAM sequence 5'-NGG-3'and unwind the DNA to check for complementarity with the gRNA. If the target DNA hybridises with the gRNA, the HNH and RuvC domains make a blunted-ends cleavage in the target, 3nt upstream the PAM. B) Representation of the dual crRNA:tracrRNA hybrid (upper panel) and of the synthetic sgRNA (lower panel). ***(Figure from Nowak et al. 2016;*** *(93)****)***.

Figure 1.11 - Representation of the gRNAs and the ribonucleoprotein Cas9:gRNA recognising its target DNA

1.8.1c - Cas9 interference process

Alone, Cas9 is in an inactive form. The binding of the gRNA induces a conformational change, which activates Cas9 and allows PAM recognition and target cleavage. The need for RNA binding to "activate" Cas9 helps reducing off-target effects. Indeed, Cas9 cannot cut the genomic DNA by itself, without being first activated by gRNA binding (94).

Once the gRNA is bound to Cas9, the first step is for the RNP to scan the genome for a PAM sequence. To do so, SpCas9 will search the genome for the 5'-NGG-3'sequence, and only upon PAM recognition, the protein will question the protospacer for gRNA complementarity (94). It is thought that this process allows a gain of time for target spotting, since, instead of questioning the genome for 20nt gRNA complementarity, Cas9 will first search for the 3nt PAM, which is a much shorter sequence and does not need an R-loop formation. On top of saving time, it also allows a decrease of offtargets, since Cas9 will not search for complementarity if there is no PAM recognition first.

PAM recognition triggers subsequent strand separation in the protospacer, allowing base pairing between the gRNA and the complementary strand in the target DNA, forming an R-loop structure. The hybridisation induces further Cas9 conformation changes, allowing the simultaneous cleavage of both the target and non-target strands, 3bp upstream the PAM sequence (Figure 1.11A). The HNH endonuclease domain of Cas9 will cut the target strand very precisely at 3nt upstream the PAM, while RuvC domain could be a bit more flexible and cut the non-target strand at 3nt or more, upstream the PAM, leading to either blunt-ends or 5' overhang ends (91, 95).

Indeed, a paper from 2018 showed that after the cells repair the double strand break created by SpCas9, there is a predictable pattern of insertions, the major one being a +1 insertion. This insertion, most of the time, corresponds to the addition of the nucleotide in the 5' overhang. The cell repaired the DSB by adding the complementary bases for the nucleotides in the single base overhangs, and then, joined the two blunt ends together (Figure 1.12; (96)).

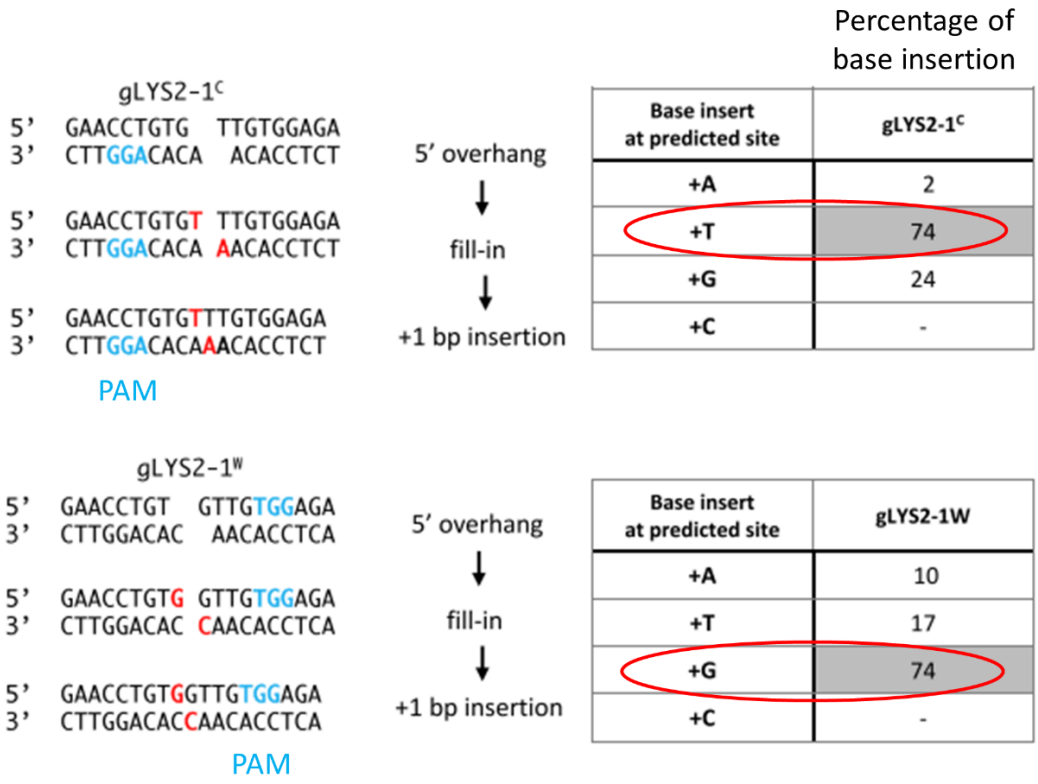


Figure 1.12: Pattern of insertion after Cas9 cutting and DSB repair. The left panel represents the expected insertion pattern if SpCas9 cut with one nucleotide 5' overhang, for two S. Cerevisiae genes. The right panel shows the percentage of base insertion observed after DSB repair. A correlation can be seen between the expected base insertion and the higher percentage of this expected base in the analyses (shown in the red circles). ***(Figure modified from Lemos et al. 2018;*** *(96))*.

Figure 1.12 - Pattern of insertion after Cas9 cutting and DSB repair

1.8.1d - Cas9 specificity

Physiologically, Cas9 is very specific and mismatches between the gRNA and the protospacer greatly reduce the cleavage efficiency of the protein (97, 98). However, studies showed that the specificity is mostly due to the 10 to 12 nucleotides just upstream the PAM, at the 3' end of the gRNA. A single mismatch in this region called "seed region", can have a detrimental effect on DNA targeting by SpCas9 (97, 98). However, although the 5' end of the gRNA is less sensitive to mismatches, more than three consecutive mismatches in that region can significantly reduce Cas9 cleavage (94, 97). Some studies showed that SpCas9 can tolerate up to 6 mismatches (97). Moreover, truncated gRNAs, smaller than 17bp, completely abolished DNA cutting, showing the importance of the 5' region for Cas9

cleavage (99). However, despite the seed sequence being essential for Cas9 activity, some studies clearly showed that the specificity is more complex and can be protospacer-dependent. Indeed, it was demonstrated that about 1 or 2 mismatches in the seed sequence could be tolerated for some gRNAs. Moreover, the localisation of mismatches that can be tolerated by SpCas9 can be different for each gRNA and its associated protospacer (97, 100).

1.8.1e - Limitations of CRISPR/Cas9 systems and solutions

Combining all this information, two main limitations of using CRISPR/Cas9 system for genome editing can be noticed. Fortunately, many scientists found many ways to overcome these inconveniences.

First, knowing that the PAM is the primary step to trigger base pairing between the protospacer and the gRNA, it is clear that the PAM sequence is a limitation for the localisation of the cleavage for genome editing. As the SpCas9 PAM sequence (5'-NGG-3') is very common (about every 8bp in the human genome), most target sites for genome editing should have a suitable PAM close by. However, some genome editing techniques require high precision. For example, homology-directed repair (HDR), which is described in section 1.9.2, is more efficient when the DSB is placed within 10-20 bp of the desired DNA modification (101). Even more precise is the base editing technique, described in the section 1.11, which requires a PAM localised around 14 to 17bp downstream the base to modify (102). Therefore, those techniques require a PAM at a very specific location, and sometimes the NGG PAM site might not be present at the site needed.

To overcome this issue, some Cas9 orthologs naturally recognises different PAMs (Table 1.1). For example, *Staphylococcus aureus* Cas9 (SaCas9) recognises the PAM 5'-NNGRRT-3', where "R" are any T or C nucleotides, and *Streptococcus thermophilus* Cas9 (St1Cas9) recognises the PAM 5'-NNAGAA-3'. Those Cas9 orthologs increase the possibilities of DNA target. Moreover, CRISPR systems have been shown to be present in about 90% of archaea and about 50% of bacteria, demonstrating that there could be many more other orthologs available in the future when those systems will be analysed. On top of the Cas9 orthologs, some engineered Cas9 nucleases were designed with altered PAM specificity (Table 1.1). For example, the variant D1135V/R1335Q/T1137R in SpCas9 (SpCas9-VQR) was engineered to recognise the PAM 5'-NGAN-3' (103).

Nuclease Name	Origin	Variant	PAM (5' -> 3')	Reference
SpCas9	*Streptococcus pyogenes*	Wild-type	NGG	Jinek 2012
SpCas9-VQR	*Streptococcus pyogenes* (Engineered)	D1135V/R1335Q/T1137R	NGAN	Kleinstiver 2015
SpCas9-EQR	*Streptococcus pyogenes* (Engineered)	D1135E/R1335/T1337R	NGNG	Kleinstiver 2015
SpCas9-VRER	*Streptococcus pyogenes* (Engineered)	D1135V/G1218R/R1335E/T1337R	NGCG	Kleinstiver 2015
NmCas9	*Neisseria meningitidis*	Wild-type	NNNNGATT	Esvelt 2013
FnCas9	*Francisella novicida*	Wild-type	NGG	Hirano 2016
FnCas9-RHA	*Francisella novicida* (Engineered)	E1369R/E1449H/R1556A	YG (Y = T or C)	Hirano 2016
SaCas9	*Staphylococcus aureus*	Wild-type	NNGRRT (R = A or G)	Kleinstiver 2015 Steinert 2015
St1Cas9	*Streptococcus thermophilus*	Wild-type	NNAGAA	Esvelt 2013 Steinert 2015

Table 1.1: Table showing examples of different Cas9 orthologs and engineered Cas9 variants, as well as their altered PAM sequences.

Table 1.1 - Example of different Cas9 orthologs and engineered Cas9 variants

The second main limitation of SpCas9 for genome editing is the mismatch tolerance outside the seed region. Indeed, this tolerance can lead to undesirable off-target effects, and some studies showed that SpCas9 can tolerate up to 6 mismatches (97). Off-target sites are DNA sequences which are homologous to the target region and can also be cleaved by SpCas9. Those cleavages can result in unwanted side effects such as cell toxicity, gene disruption, or genome instability.

To overcome this issue, new synthetic nucleases were designed such as the "**e**nhanced specificity" SpCas9 (eSpCas9) (104), the "**H**igh-**F**idelity" SpCas9 (SpCas9-HF1) (105), the "**Hyp**er-**a**ccurate" SpCas9 variant (HypaCas9) (106) and the "**evo**lved" SpCas9 (evoCas9) (107). Each of them was shown to have significantly lower off-target effects while still demonstrating similar on-target effects as the WT SpCas9. For example, evoCas9 showed 79-fold specificity improvement, while keeping 90% of the on-target editing efficiency (107).

Moreover, another way to limit off-targets is the use of the many available computational tools. Those programmes are free and available online to help scientists designing the most optimal gRNAs for their experiments. They will give all the available gRNAs in the region of interest and give a score for on-target and off-target effects.

The scoring is based on the presence of similar sequences in the genome of the organism of interest, the number of mismatches between the gRNA and the potential targets, their location in the gRNA and their identity. Those programmes are great tools to design gRNAs with a maximum reduction of off-target effects.

Although many tools are available to decrease off-target effects, whenever a scientist is manipulating the human genome, off-target effects should be assessed for safety. Therefore, many methods are available to detect off-targets. Those methods are listed and described in this review from Manghwar et al. in 2020 (108).

1.8.2 - CRISPR/CAS12a

Cas12a, originally described as Cpf1 (**C**RISPR from ***p****revotella* and ***f****rancisella***1**), is about 300 amino acids shorter than SpCas9 with about 1300 amino acids. Like Cas9, Cas12a is a bi-lobed protein with a recognition (REC) lobe and nuclease (NUC) lobe. Moreover, the heteroduplex crRNA and protospacer's target-strand DNA is accommodated in the channel between those two lobes. Finally, like Cas9, Cas12a possesses a RuvC-like endonuclease domain (109).

1.8.2a - Cas9 vs Cas12a

Even with those many structural similarities, the Cas12a protein shares no similar sequences to Cas9, except for the RuvC-like domain. There are three main differences between the two nucleases (Figure 1.13):

1) Cas12a requires a crRNA, without the involvement of a tracrRNA (109, 110). The Cas12a crRNA is about 42bp long, with 19nt coming from the repeat sequence and about 23nt corresponding to the spacer (109). The Cas12a crRNA is smaller compared to Cas9 (42nt crRNA + 89nt tracrRNA (91)), and has a simpler secondary structure, consisting on a single stem loop, which makes it cheaper and easier to synthesise (109, 110). Moreover, those characteristics confer advantages for transfection in cells or encapsidation in viral vectors, which can be size limiting.

2) The PAM requirement for Cas12a is a T-rich sequence located about 18bp upstream of the cut site on the non-target strand (Figure 1.13; (110)). Two advantages can be associated with those characteristics. First, both the T-rich PAM from Cas12a and the G-rich PAM from Cas9 together, expand the range of site target. Second, because Cas12a cuts far from the PAM, at the end of the protospacer, there would be an advantage for homology directed repair (HDR)-based genome editing. Indeed, with Cas9, if **N**on-**H**omologous **E**nd **J**oining (NHEJ) repair happens, the insertions and deletions (indels) formed can disrupt the PAM and/or the seed sequence, making Cas9 unable to cut a second time. However, using Cas12a, if NHEJ creates indels, most of the time, the PAM will be still intact as most of the protospacer. Therefore, there will be a higher chance that Cas12a could cut again, giving the cell another opportunity to perform HDR repair (110).

3) Cas12a cleave the protospacer usually after the 18th base on the non-target strand and after the 23rd base on the target strand (110). The double strand break generated results therefore in staggered ends with five nucleotides 5' overhangs (Figure 1.13). This characteristic is advantageous for NHEJ repair, since one can design an insert with the matching overhangs, facilitating insertion in the correct orientation and increasing efficiency, compared to an insert with blunt ends that can integrate in both orientations (110). Importantly, Cas12a possesses only a single RuvC-like endonuclease domain, which prevents the design of a Cas12a nickase (109).

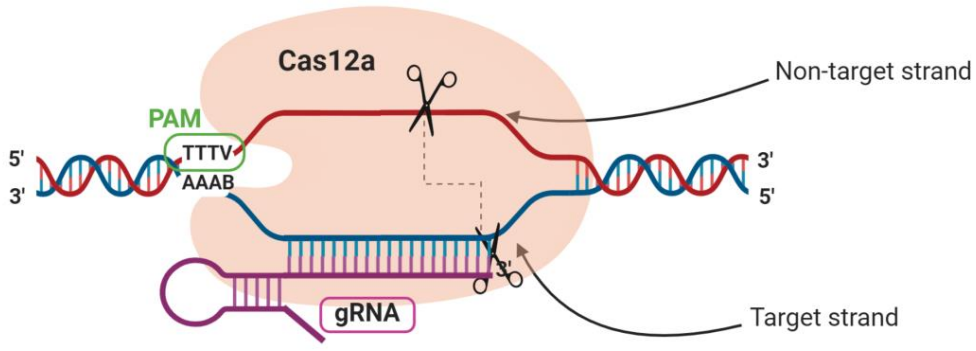


Figure 1.13: Schematic representation of protospacer recognition by AsCas12a. The protein recognises the 5'-TTTV-3' PAM sequence, leading to opening of the target DNA duplex and binding of the gRNA. Upon base pairing, Cas12a induces a staggered cut at about 18nt on the non-target strand, and about 23nt on the target strand.

Figure 1.13 - Protospacer recognition by AsCas12a

1.8.2b - Cas12a protein structure

As described previously, Cas12a has a bi-lobed architecture composed of the REC lobe, containing REC1 and REC2, and the NUC lobe, containing the PAM-interacting (PI), wedge (WED), bridge helix (BH), RuvC and Nuc domains (Figure 1.14; (109)). Structural studies showed that, in the absence of the crRNA, the REC2 domain occludes the RuvC catalytic site, which prevent cleavage (111). Without activation of Cas12a with a crRNA, the enzyme is therefore in an inactive conformation.

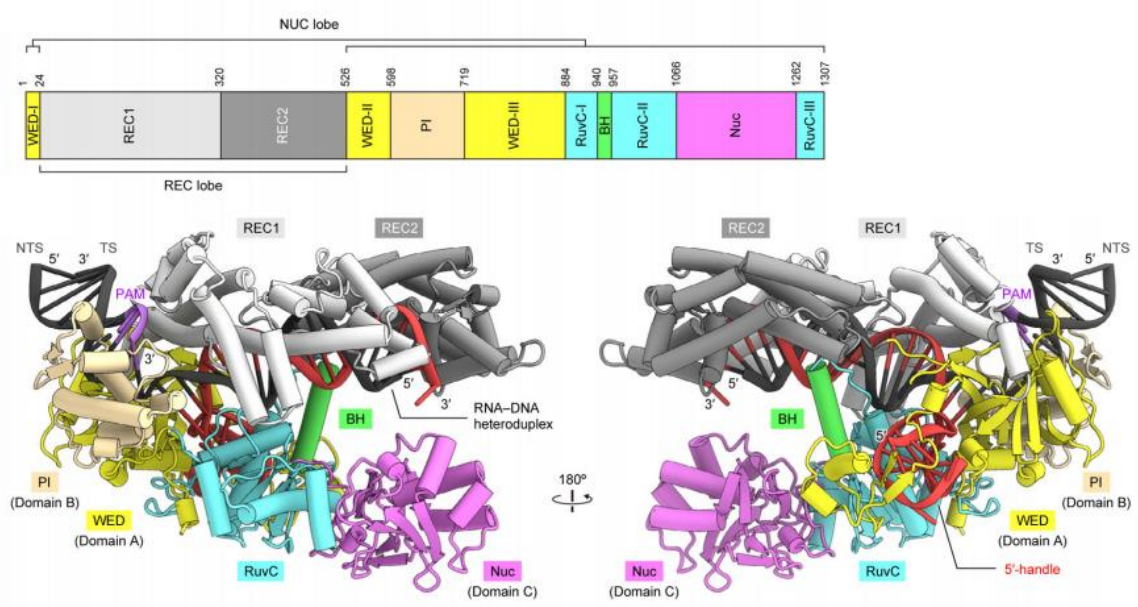


Figure 1.14: Lachnospiraceae bacterium Cpf1 (LbCpf1) protein structure A) Domain organisation of the LbCpf1protein. B) Schematic structure of the LbCpf1 protein with its crRNA. The scheme shows the localisation of every domain relative to the crRNA. ***(Image adapted from Yamano et al. 2015*** (109)**).**

Figure 1.14 - Lachnospiraceae bacterium Cpf1 (LbCpf1) protein structure

1.8.2c - Cas12a target recognition and cleavage

Like Cas9, the first step for target recognition is the detection of the PAM sequence. In the case of *Acidaminococcus sp.* Cas12a (AsCas12a), which is one of the most efficient Cas12a nucleases in mammalian cells (112), the PAM sequence is 5'-TTTV-3', where "V" corresponds to any nucleotide except a thymine (110). Upon PAM recognition, there is strand invasion of the crRNA, which progressively binds the target-strand of the protospacer, leading to DNA duplex unwinding and R-loop formation (111).

The hybridisation of the crRNA with the target strand creates a Cas12a conformation change, which catalytically activates the protein for cleavage. Indeed, upon hybridisation, the REC lobe undergoes a conformational change with the REC1 domain rotating about 33° relative to the NUC lobe, so it gets closer to the latter (111). In the meantime, the REC2 domain rotates 46° relative to REC1, to bind the PAM-distal end of the crRNA:protospacer target strand heteroduplex. Those conformational changes expose the RuvC nuclease site for DNA cutting (111). Unlike Cas9, Cas12a possesses only the RuvC endonuclease domain, and does not have an HNH endonuclease domain (109). Therefore, in Cas12a, the RuvC domain is the only one responsible for the cuts of both the target and non-target DNA strands (109). Mutations in the RuvC domain completely inactivate Cas12a cleavage (109). It is supposed that Cas12a cuts both strands sequentially, starting with the non-target strand, followed by cleavage of the target strand (109, 111). First, the nucleotides 7 to 13, from the PAM site, in the non-target strand, interact electrostatically with a positively charged groove close to the RuvC domain (111). This interaction is thought to contribute to the stabilisation of this displaced non-target ssDNA in the R-loop structure, and also contribute to the positioning of this non-target strand into the RuvC active site for cleavage. Studies blocking the cleavage of one or the other strand showed that the cleavage of the non-target strand was a prerequisite for the target strand cleavage (111). A theory is that the first cleavage enables a local unwinding of the DNA duplex at the PAM-distal location, moving the target strand into the RuvC cleavage site (111). It was thought that the Nuc domain plays a role in the second strand cleavage. The idea would be that Nuc domain places the target strand into the RuvC catalytic site for subsequent cleavage (113). Upon cleavage of both strands, the PAM distal DNA duplex is released, while Cas12a stay fixed on the PAM proximal side of the DSB. Therefore, crRNA:Cas12a RNP is still bound to the target strand, keeping Cas12a in an active conformation for DNA cutting (111).

Recently, Cas12a was also shown to cleave ssDNA substrates, sequence-independently. This activity has been called trans-cleavage activity (111). Basically, the catalytically active Cas12a, which is still attached to the protospacer and has its RuvC domain exposed, is able to cut any ssDNA that passes through its catalytic site (Figure 1.15; (111)). However, the exact mechanism for trans-cleavage remains to be determined. For trans-cleavage activity, the binding of Cas12a to the protospacer, the double-stranded cleavage and the protospacer ends dissociation are a prerequisite (111).

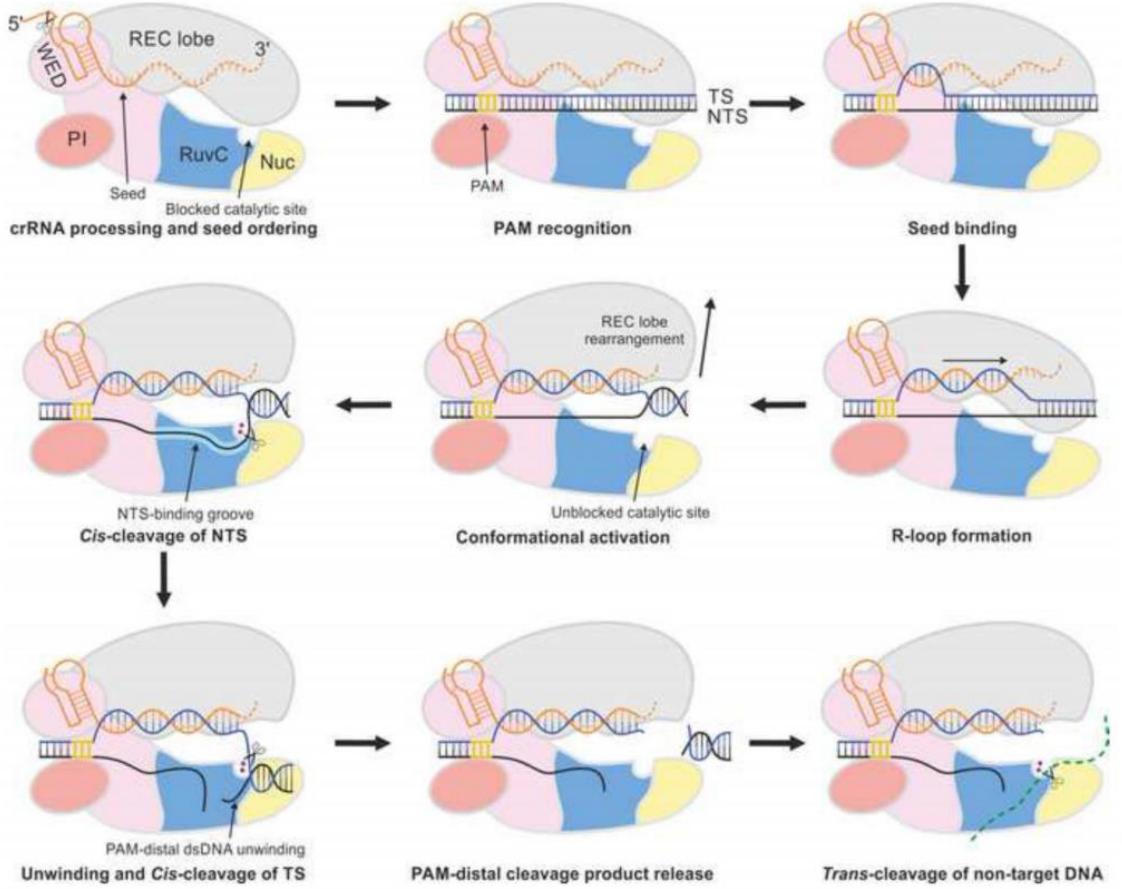


Figure 1.15: Schematic representation of the model proposed by Daan Swarts and Martin Jinek, for Cas12a cis and trans-cleavages. First, the protein recognises the PAM sequence from the protospacer, leading to the unwinding of the protospacer DNA duplex. The crRNA binds to the target strand, forming an R-loop structure. The heteroduplex crRNA:protospacer DNA formation leads to a conformational change of the Cas12a protein, and activation of the RuvC cleavage capacity. The non-target strand is cleaved by RuvC, followed by cleavage of the target-strand. Once the DSB is made, the PAM distal end is freed, and Cas12a stays attached to the PAM proximal end, in an active state, potentially cleaving any ssDNA passing by the RuvC site, which stays exposed. ***(Figure from Swarts and Jinek. 2019;*** *(111))*.

Figure 1.15 - Model for Cas12a cis and trans-cleavages

1.8.2d - High specificity of Cas12a

Cas12a was shown to be highly specific in human cells with the nuclease tolerating only 1 or 2 mismatches. Studies showed that two adjacent mismatches at positions 1 to 18 in the crRNA:protospacer heteroduplex, inhibit Cas12a cleavage. However, mutations after position 18 are tolerated (112, 114). Some tolerances were observed for mismatches at positions 1, 8 and 9, but overall, Cas12a is sensitive to mismatches up to position 18.

Moreover, the seed region of Cas12a has been shown to be the five first nucleotides at the 5' end of the crRNA, in the heteroduplex (110). Studies also showed lower off-target for Cas12a compared to Cas9, potentially making the Cas12a nuclease a safer protein to use for genome editing (112).

1.8.2e - Efficiency of Cas12a

Considering the AsCas12a gene editing efficiency, studies showed that, overall, AsCas12a was less efficient than SpCas9, for both NHEJ (112, 115) or HDR (115) repair pathways. However, AsCas12a was equally or more efficient than other Cas9 orthologs such as SaCas9, StCas9 or NmCas9 (116).

Concerning NHEJ, Kim et al. showed that, in average, for 10 different chromosomal target sites, in human cells, SpCas9 was editing at a frequency of 32 ± 4%, while AsCas12a was editing at a frequency of 20 ± 5% (112). Wang et al. showed that, in average, for 61 genomic loci in HEK293T cells, SpCas9 exhibited the highest cleavage efficiency (~48%) while AsCas12a had a cleavage efficiency of about 30% (115).

Concerning HDR, Wang et al. compared the HDR efficiencies for the nucleases, in two different HEK293T cells. They showed that with optimal ssODN donors, SpCas9 was editing, in average, at a frequency of 10% on the *CACNA1D* gene and 20% on the *PPP1R12C* gene. AsCas12a exhibited 10% and 8% respectively. Using a linearised plasmid donor, SpCas9 was editing, in average, at a frequency of 7% on the *CLTA* locus and 7% on the *GLUL* locus. AsCas12a was editing, on average, at a frequency of 8% and 1% respectively.

1.8.2f - Limitations of Cas12a

There are three main disadvantages for the use of CRISPR/Cas12a:

First, as for Cas9, the PAM could be a sequence limitation for genome editing techniques that require high specificity. The group from Feng Zheng's lab, in 2017, designed two Cas12a variants S542R/K607R and S542R/K548V/N552R, which recognise 5'-TYCV-3' (cut about every 96bp) and 5'-TATV-3' (cut about every 48bp) PAMs respectively ("Y" being a Cytidine or a Thymine) (117). Those new PAMs all together, would allow the protein to cleave about every 11bp (117). SpCas9, AsCas12a and their orthologs or synthetic variants, all together, greatly increase the targeting ranges for genome editing.

Second, it was observed that the efficiency of AsCas12a seems to be lower than the one from SpCas9. However, one thing to notice is that Cas12a is a very recent protein, therefore, while Cas9 has been synthetically modified to improve its efficiency, Cas12a had not been yet. However, the last few months, a new AsCas12a from IDT (Alt-R A.s.Cas12a (Cpf1) *Ultra* Nuclease, IDT) was made available for research, and had showed to greatly increase the editing efficiency compared to the previous AsCas12a (118).

However, to the best of my knowledge, no comparisons with Cas9 have been done yet. Importantly, a paper from 2019, made a comparison between a SpCas9 and an AsCas12a from the company Aldevron, and demonstrated that Cas12a was outperforming Cas9 (119). In this paper, the group designed a nanoformulation for transfection in haematopoietic stem and progenitor cells without toxicity. They used this nanoformulation to transfect both Cas9 and AsCas12a, targeting the *CCR5* gene. The results showed that AsCas12a was outperforming Cas9 for both NHEJ repair pathway (18% vs 10% respectively) and HDR repair pathway (14% vs 4% respectively; (119)).

An idea could be that the Aldevron Cas12a could have been optimised to have an increased efficiency compared to the other Cas12a enzymes used in previous studies. Therefore, the use of Cas12a still displays a very high potential for genome editing.

Finally, the third limitation for Cas12a is its trans-cleavage. Indeed, section 1.8.2c showed that Cas12a can cut ssDNAs sequence-independently. Importantly, most of the recent HDR studies performed with Cas9 and Cas12a, used **s**ingle-**s**tranded oligodeoxyribonucleotides (ssODNs) as donor templates. Therefore, it could be possible that Cas12a degrades the ssODN donor template, thus decreasing the HDR efficiency. It could explain the lower efficiency of Cas12a compared to Cas9 for HDR, seen in previous papers (115).

1.9 DOUBLE STRAND BREAK REPAIR MECHANISMS

Following double stranded breaks made by the nucleases, cells possess several mechanisms that can repair them. The pathway used to correct the double strand break will define the outcome of the editing. There are two main pathways for repair: Non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Figure 1.16).

The non-homologous end joining pathway is the fastest repair pathway and is used throughout the cell cycle in mammals. Indeed, studies showed that NHEJ can be completed in about 30min (120). This pathway mostly consists on direct ligation of the two ends flanking the break, using little or no sequence homology. NHEJ pathway is usually very precise but it could happen that short insertions or deletions (indels) of some genetic information occur. Those indels formation are the reason why NHEJ is usually considered a mutagenic repair pathway. For gene editing, NHEJ is usually used for knock-out of genes, by using indels to make a gene frameshift (Figure 1.16).

Regarding homology-directed repair (HDR), this repair pathway contains several subpathways. For the sake of being concise, this introduction will focus on the two subpathways the most relevant for genome editing: **S**ynthesis-**D**ependent **S**trand **A**nnealing (SDSA) and canonical **H**omologous **R**ecombination (HR) (also called canonical **D**NA **D**ouble-**S**trand **B**reak **R**epair, DSBR; (121)). HDR, as its name indicates, uses a template with homologies, to copy and insert the missing information into the genome.

Physiologically, the template is the sister chromatid, and consequently, this process is mostly active during the steps of the cell cycle when the sister chromatid is present (phase S and phase G2; (122)). Due to the requirement of a template for repair, the use of HDR pathway is highly accurate and allows complete restoration of the genetic information. Regarding genome editing, the introduction of an exogenous donor template allows precise genome modifications, and introduction of a desired sequence. Because HDR is mechanistically more complicated and necessitates a larger number of enzymes, this pathway slow and studies showed that HR can take more than 7 hours to be completed (120).

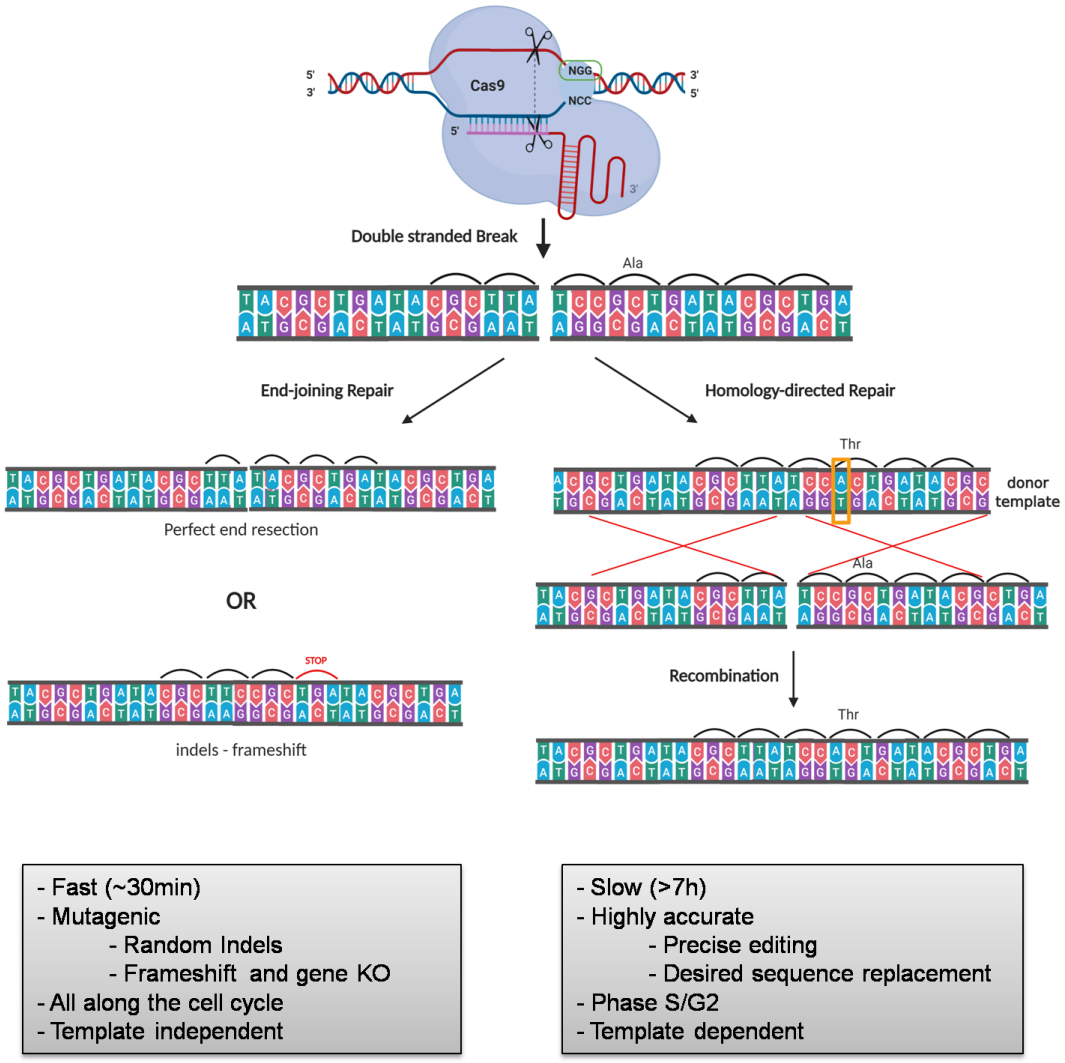


Figure 1.16: Figure representing the two main repair pathways after DSB created by Cas9. On the left, NHEJ repair pathway can lead to a frameshift through indels formation. The frameshift mainly leads to knock-out of genes. On the right, HDR uses a donor template to trigger recombination. The information from the template is copied to the break, resulting in a very accurate repair.

Figure 1.16 - The two main repair pathways (NHEJ and HDR), after DSB made by Cas9.

1.9.1 - NON-HOMOLOGOUS END-JOINING REPAIR PATHWAY

NHEJ is the most common DSB repair pathway in humans. The main purpose is to make the two broken ends compatible, sometimes with some end-processing, in order to bind them together. If the break is blunt, the two ends can ligate together without any processing. However, if there are incompatible overhangs, end processing will be achieved in order to expose a short homologous sequence of ~2 to 4nt, enabling ends binding, DNA synthesis and ligation.

End repair using NHEJ can be divided into 5 stages: 1) The free ends are recognised by Ku heterodimer, 2) The two ends are kept in proximity through end-bridging, 3) if needed, the DNA ends are processed to produce compatible ends, 4) the ends are ligated together and 5) Ku is removed from the repaired DNA (123).

In eukaryotes Ku is a heterodimer composed of the two subunits Ku70 and Ku80. Ku are very abundant with each cell possessing about $4x10^5$ copies of the protein. Moreover, they have a very high affinity for free DNA ends (124, 125). Those characteristics allow them to bind the DNA ends only within 5 seconds after the damage. Therefore, upon DNA break, Ku are the first proteins binding to the DNA ends at the site of the damage, making this pathway the fastest for DSB repair. Ku binding protects DNA ends from degradation and allows recruitment for other NHEJ proteins. NHEJ repair can be precise, but it becomes inexact when the precise repair is not an option (126). For each type of break (blunt end, or ssDNA overhang), there will be a specific NHEJ process.

1.9.1a - NHEJ for blunt ends breaks

Regarding blunt ends, such as most of the cut made by Cas9 nucleases, the ends containing Ku proteins attract the **X**-ray **R**epair **C**ross-**C**omplementing protein **4** (XRCC4) - DNA ligase IV complex. The complex is composed of two XRCC4 for one DNA ligase IV, which forms a bridge between the two DNA ends. The ligase leads to blunt-ends direct ligation, without any strand processing. This pathway usually results in perfect resection (Figure 1.17; (127)).

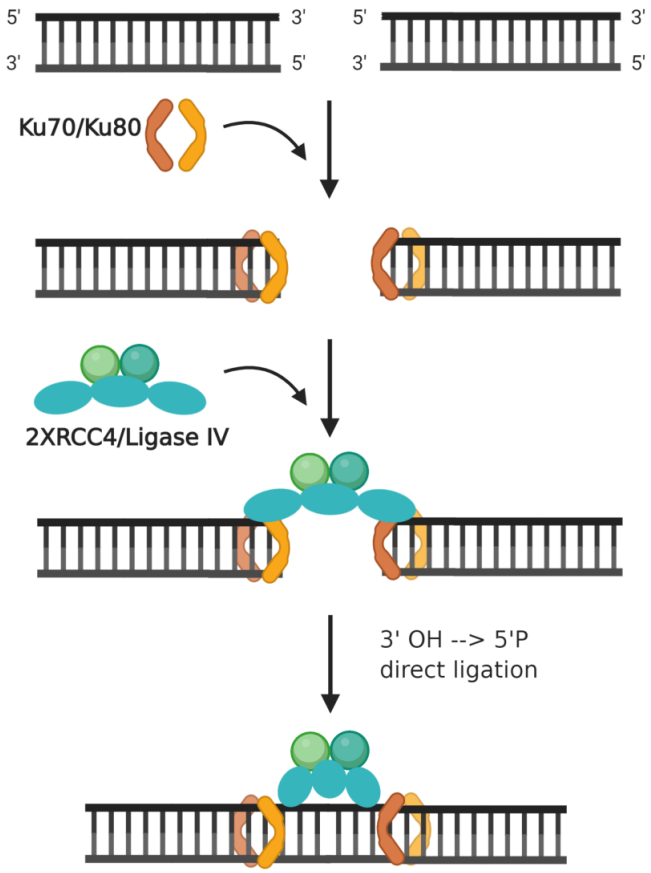


Figure 1.17: Graphic representation of NHEJ repair pathway for blunt ends, after DSB. First, the Ku complex binds to each DNA end, recruiting the 2XRCC4/Ligase IV complex to the place of the damage. The ligase binds the ends together, resulting in perfect resection of the break.

Figure 1.17 - NHEJ repair pathway for blunt ends

1.9.1b - NHEJ for breaks with overhangs

In the cases where the damage creates ends with non-compatibles overhangs (mismatches or chemical modifications), a resection takes place in order to create blunt ends or expose 3'single strand DNA (ssDNA) on each side of the break (128). Those 3' overhangs will allow a search for short homologies ($\leq$4nt) and enable end-joining (128). When resection is required for NHEJ, a complex composed of the DNA-dependent protein kinase catalytic subunit (DNA-PK_{CS}) and the Artemis nuclease is recruited to the break (128). Upon binding to the DNA ends, DNA-PK_{CS} undergoes autophosphorylation, and Artemis' phosphorylation, which activates the latter (128, 129). Artemis, then, gains the ability to cleave DNA at the junction between ssDNA and dsDNA (128).

Artemis has an intrinsic 5’ exonuclease activity on ssDNA. However, when complexed with autophosphorylated DNA-PK$_{CS,}$ the nuclease gains an endonuclease activity (128, 130). The complex Artemis/DNA-PK$_{CS}$ binds a 4nt stretch of ssDNA at the ssDNA/dsDNA junction, and cuts on the 3’ side of those 4nt. As a consequence, a 5’ overhang will lead to a nick directly at the junction, generating a blunt end, while a 3’ overhang will lead to a shorter overhang of about 4nt (Figure 1.18; (128, 130)).

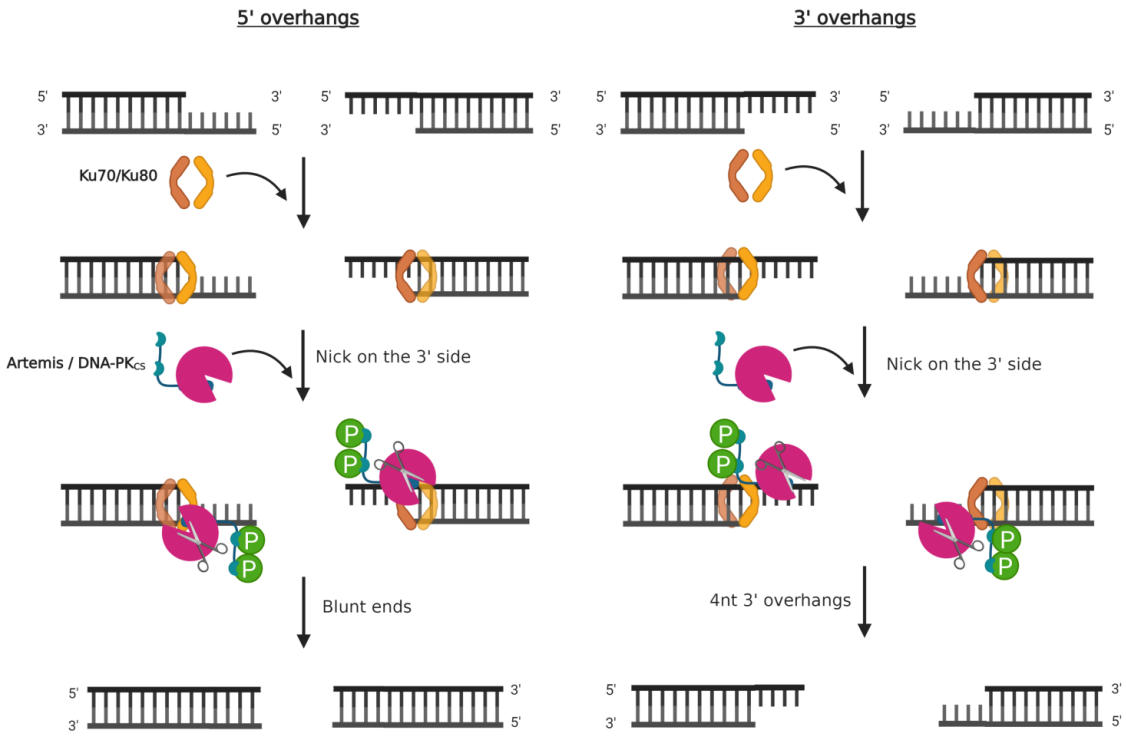


Figure 1.18: Schematic representation of Artemis initial end-resection, with regard to the type of overhang available. A 5’ overhang will lead to Artemis cleaving the DNA at the ssDNA/dsDNA junction, while a 3’overhang will lead to Artemis cleaving the DNA 4nt from the junction on the ssDNA.

Figure 1.18 - Artemis initial end-resection depending on the type of overhang available

Regarding 5’overhangs, such as ends resulting from Cas12a cleavage, the processing by Artemis results in compatible blunt ends, which can be ligated with the ligase IV. Conversely, 3’ overhangs processed by Artemis result in shorter 3’ overhangs, and Artemis will further excise the 5’ DNA strand, resecting into the dsDNA, to generate longer stretches of 3’ssDNA and expose microhomologies (128).

Those homologies are shared by chance between the two ends. The short base pairing between the two ends will lead to cleavage of the extra non-compatible ssDNA located before the homology, DNA synthesis of the gaps, and ligation of the ends. For example, if one end contains a 5′---CTGGTTCA*3′ and the other end is 5′*tcgttagt---3′ (where (-) represents the remaining DNA and the (*) represents the break site), the "GTT" represent the homology in both sequences. Therefore the 5′-GTT-3′ from the first end, can bind to the complementary 3′-caa-5′ from the other end. After cutting the excess 3' DNA ("CA" and "tc"), the gaps can be filled with a polymerase and the ends can be ligated. The final NHEJ repaired sequence would be 5′---CTG**GTT**agt---3′ (the bold letters being the homology recombination point; Figure 1.19; (126)).

The final products after NHEJ repair of incompatible overhangs mainly result in short deletions around the site of damage.

DNA Polymerases μ and λ (pol μ, pol λ) both have a role in NHEJ repair pathway. Pol μ tends to act more in a template-independent manner, while pol λ has primarily a template-dependent activity. Pol λ is mostly responsible for the fill-in of the gap after microhomology pairing. Pol μ, on the contrary, is responsible of the "insertions" created after NHEJ. Indeed, if no microhomologies are present in the ssDNA overhangs, pol μ can add few nucleotides at the end of the overhangs, in a template-independent manner, in order to create terminal microhomologies for pairing and ligation. This results in nucleotides template-independent insertions at the break junction, after repair by NHEJ.

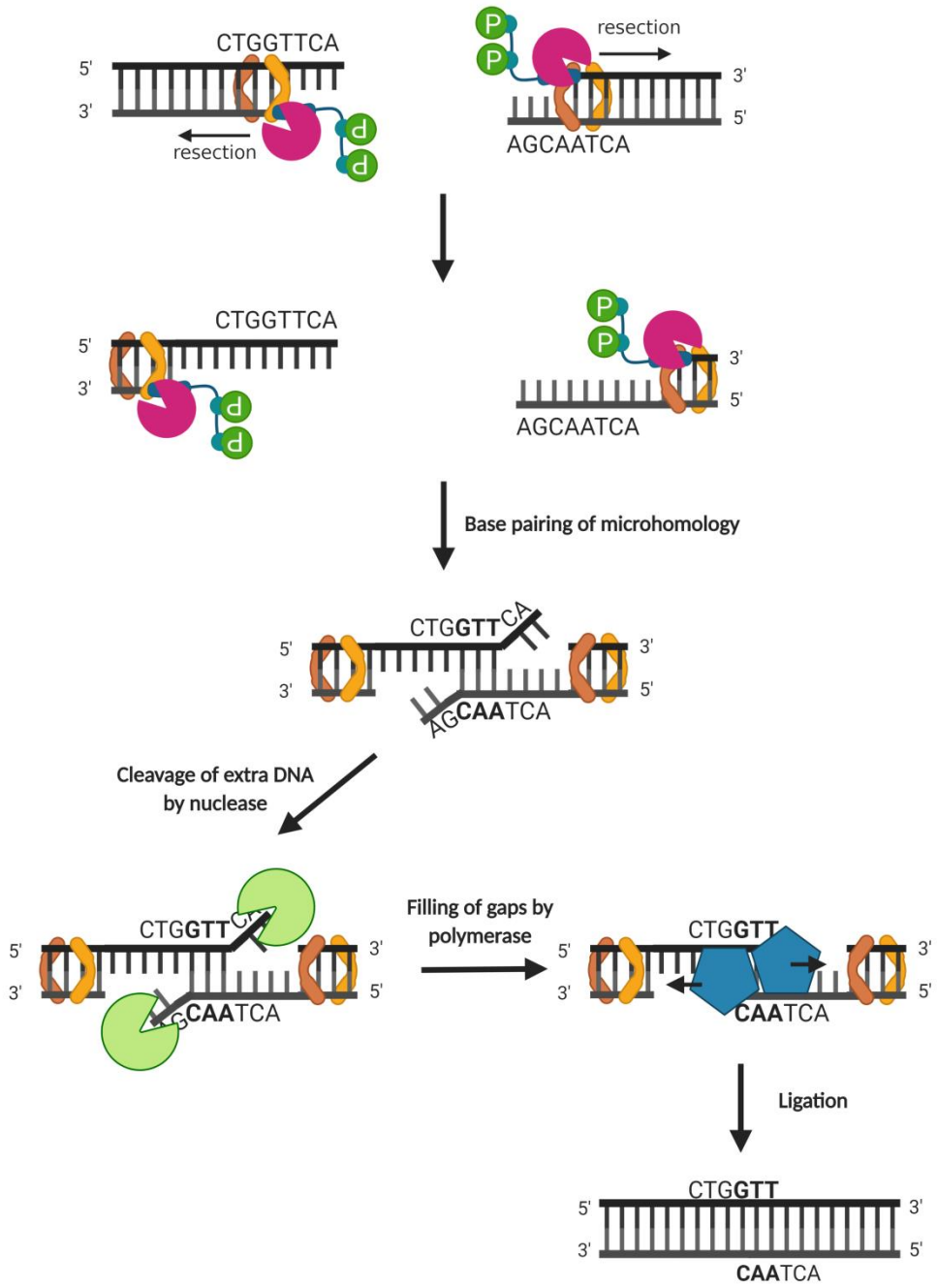


Figure 1.19: NHEJ repair of incompatible 3' overhangs. Artemis/DNA-PK$_{CS}$ complex further resects the DNA by cleaving into the dsDNA. This process reveals longer stretches of ssDNA that can have random microhomologies of about 2 to 4nt. The two ssDNAs hybridise at the homology site, and the extra DNA is cleaved out. The gaps are filled by polymerases and the ends are ligated together. The whole process results in short deletions of the DNA after repair.

Figure 1.19 - NHEJ repair of incompatible 3' overhangs

1.9.2 - HOMOLOGY-DIRECTED REPAIR (HDR) PATHWAY

DNA recombination using HDR pathway can be divided into 4 steps: 1) the "**initiation**" consists of end-resection, to create very long ssDNA stretches, 2) the "**presynaptic phase**" consists of the formation of the nucleoprotein filament ssDNA/RAD51, and the search for the homologous template DNA duplex. 3) The "**postsynaptic phase**" consists of the strand invasion of the template, formation of the Displacement loop (D-loop), and replication of the template. Finally, 4) the separation between the template and the repaired genomic DNA consists of the **removal of the Holliday junctions by dissolution or resolution** (131).

1.9.2a - Initiation

The initiation for homologous recombination is characterised by the long processing of the broken DNA ends, producing ssDNA long enough for homologous DNA recognition. The resection of the 5' terminal ends occurs in two main steps:

The initial resection is called **short-range DNA resection**. This process necessitates the MRE11-RAD50-NBS1 (MRN) complex, together with CtIP protein (122, 132). The MRN complex possesses an exonuclease 3'→5' activity. However, because HDR uses long 3' DNA stretches, the assumption is that the complex cuts the 5' end endonucleolytically, distal from the cut site, and then, resect exonucleolytically 3'→5' towards the break (121). This initial resection can create up to 300nt of 3'ssDNAs and is believed to have a role for bypassing proteins or modifications localised on the DNA ends at the break site (121).

The second step following this initial resection is called **the long-range DNA resection**. This step is characterised by a cleavage of the DNA in a 5'→3' direction. The cleavage creates long 3' ssDNAs, which can extend to several kilobases in length (121). Following the short-range resection, the MRN complex that is still on the DNA, enables the recruitment of some 5'→3' nucleases. Two main nucleases are thought to play a role in this second step: EXO1 or/and DNA2 (121, 132). EXO1 is an exonuclease which specifically degrades 5'-terminated ends within a dsDNA (133), while DNA2 can only degrade 5'-terminated ssDNA (132, 134). Consequently, DNA2 needs a helicase cofactor. The Werner syndrome ATP-dependent (WRN) and/or Bloom syndrome (BML) helicases are assumed to be DNA2 cofactors (135). The helicase will separate the two strands from the dsDNA, making a ssDNA substrate available for DNA2 (Figure 1.20; (121, 135)).

Those long 3'ssDNA stretches are coated with replication proteins A (RPA), which are placed all over the ssDNA to protect it from nucleases or to prevent secondary structure formation.

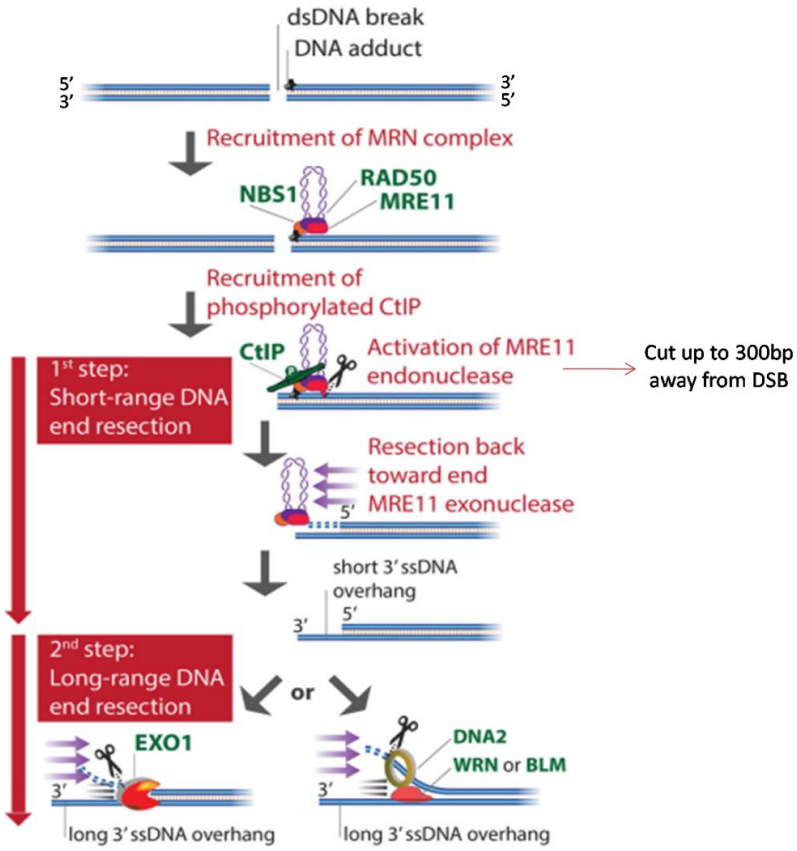


Figure 1.20: Graphic representation of the initiation step for HDR. The MRN complex is recruited at the break, followed by CtIP, which activates MRE11 for DNA cleavage. MRE11 makes a nick endonucleolytically into the dsDNA and then resects exonucleolytically towards the break. After this short-range resection, EXO1 or DNA2 further resects the 5'ssDNA to make longer 3'ssDNA overhangs. ***(Figure adapted from Ranjha. 2018;*** *(121)****).***

Figure 1.20 - Initiation step for HDR

Noticeably, CtIP, which is the MRN cofactor, needs to be phosphorylated to promote the initial short-range resection. Some of those phosphorylations are under the control of **C**yclin-**D**ependent **K**inases (CDK), which control the cell-cycle (122).

Therefore, CDK phosphorylation controls the resection and consequently, the start of the HDR pathway (122). This control allows HDR to be initiated only during the S and G2 phases, when the sister chromatids are available as a template for repair.

1.9.2b - Presynaptic phase

At this step of the HDR pathway, the two sides of the DNA break are composed of long stretches of 3'ssDNA protected by RPA proteins. To start the search for homologous DNA, the first step is the replacement of those RPA with RAD51 proteins, step mediated by the protein BRCA2 (Figure 1.21; (136)). The presence of RAD51 on the ssDNA stretches forms nucleoprotein filaments called the "presynaptic complex" because it is formed before the "synaptic" step, which corresponds to the DNA pairing between the RAD51 nucleoprotein filament and the dsDNA template. The formation of the RAD51 nucleoprotein filament induces the next step which is the search for the dsDNA template and pairing (121). The nucleoprotein filament has a high mobility, which allows it to sample a large volume of chromosomal space (137, 138). The exact mechanism of homology search is not well defined, however, it is thought that the presynaptic complex randomly makes short contacts with different DNA duplexes (139, 140). Only when the presynaptic complex recognises a sequence with at least 7nt homologies, does the complex stabilise and search for additional homologies around. A study in yeast showed that the repair efficiency was decreased when the template sequence had mismatches about every 8 nucleotides (141). The importance of this level of mismatch is used for HDR experiments in Chapter 6.

1.9.2c - Postsynaptic phase

Upon homology identification, the presynaptic complex invades the template DNA duplex, separating the two strands and forming a D-loop. This D-loop allows the 3'ssDNA to bind its complementary sequence in the template by Watson-Crick base pairing. The strand of the D-loop that is now single-stranded is being coated with RPA protein to prevent reversion of the D-loop (Figure 1.21; (142)). Moreover, a RAD54 translocase protein further stabilises the D-loop by removing RAD51 from the 3'ssDNA strand (143). RAD51 removal facilitates HDR by enabling DNA synthesis with the invader 3'ssDNA as a primer for synthesis (121). The DNA synthesis leads to recovery of the genetic information at the break site, using the homologous DNA as template (Figure 1.21). Once the 3' ssDNA invaded the template DNA duplex and copied the information, this is where the hub between the two pathways: synthesis-dependent strand annealing (SDSA) and the canonical DNA double-strand break repair (DSBR), mentioned in section 1.9, takes place.

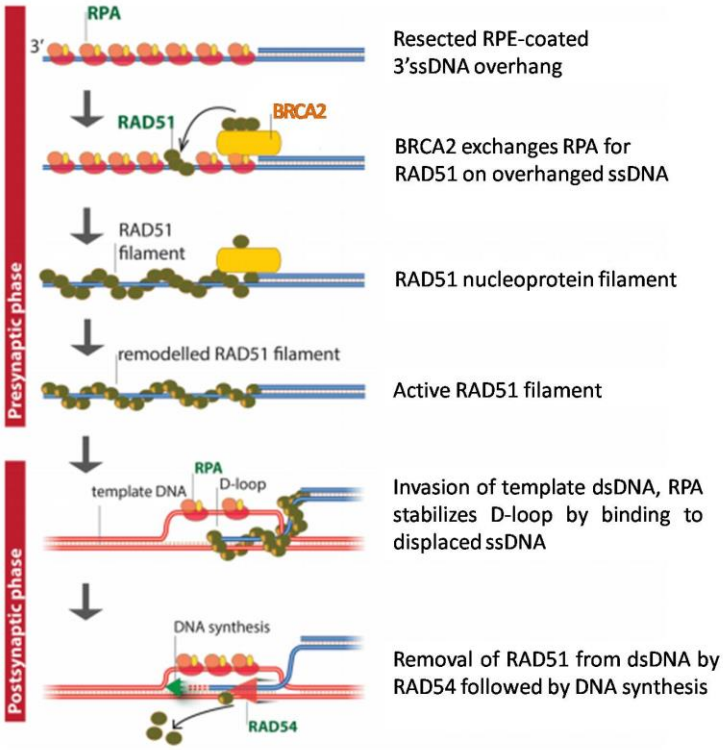


Figure 1.21: Graphic representation of the presynaptic and postsynaptic phases. The presynaptic phase corresponds to the replacement of RPA with RAD51, forming the nucleoprotein filament. The filament will search for a homologous sequence, leading to the postsynaptic phase, which corresponds to the base pairing of the 3'ssDNA with the template. The RAD51 nucleoprotein filament will invade the template DNA duplex and hybridise with the complementary strand. The second strand is displaced, creating a D-loop structure. The pairing of the 3'ssDNA is followed by replication of the template to restore the missing information. ***(Adapted from Ranjha et al 2018;*** *(121)).*

Figure 1.21 - Presynaptic and postsynaptic phases for HDR

At this point of the HDR process, a control point will determine the pathway that will be used. The balance between each pathway depends on the control of the D-loop stability. Basically, if the D-loop is disrupted, the pathway will lead to SDSA. However, if the replication is allowed to continue, then the D-loop will be stabilised and lead to DSBR (HR) pathway (121).

In SDSA, the replicated invading strand detaches from the template and anneals with the 3'ssDNA from the other side of the DSB. The SDSA repair is completed after DNA synthesis to fills the gaps and DNA strands ligation (Figure 1.22a).

Some DNA helicases such as BLM, the ATP-dependent DNA helicase Q1 (RECQ1), and the Regulator of Telomere Elongation helicase 1 (RTEL1), have been shown to play a role in disrupting the D-loop to promote SDSA (121, 144-146).

In DSBR (or homologous recombination, HR), the D-loop is more stable, making the displaced ssDNA template from the D-loop available as a substrate for the second resected 3'ssDNA end (121). This process is called "second DNA end capture" and lead to a "double" D-loop structure (Figure 1.22b; (147)). The second end capture is followed by DNA synthesis and ligation of the DNA ends. This ligation gives rise to two physical DNA junctions between the genomic DNA and the template DNA, which are called "**d**ouble **H**olliday **J**unctions" (dHJs) (Figure 1.22b; (148)). Those HJs need to be removed in order to separate the template DNA from the genomic DNA (Figure 1.22b; (121)).

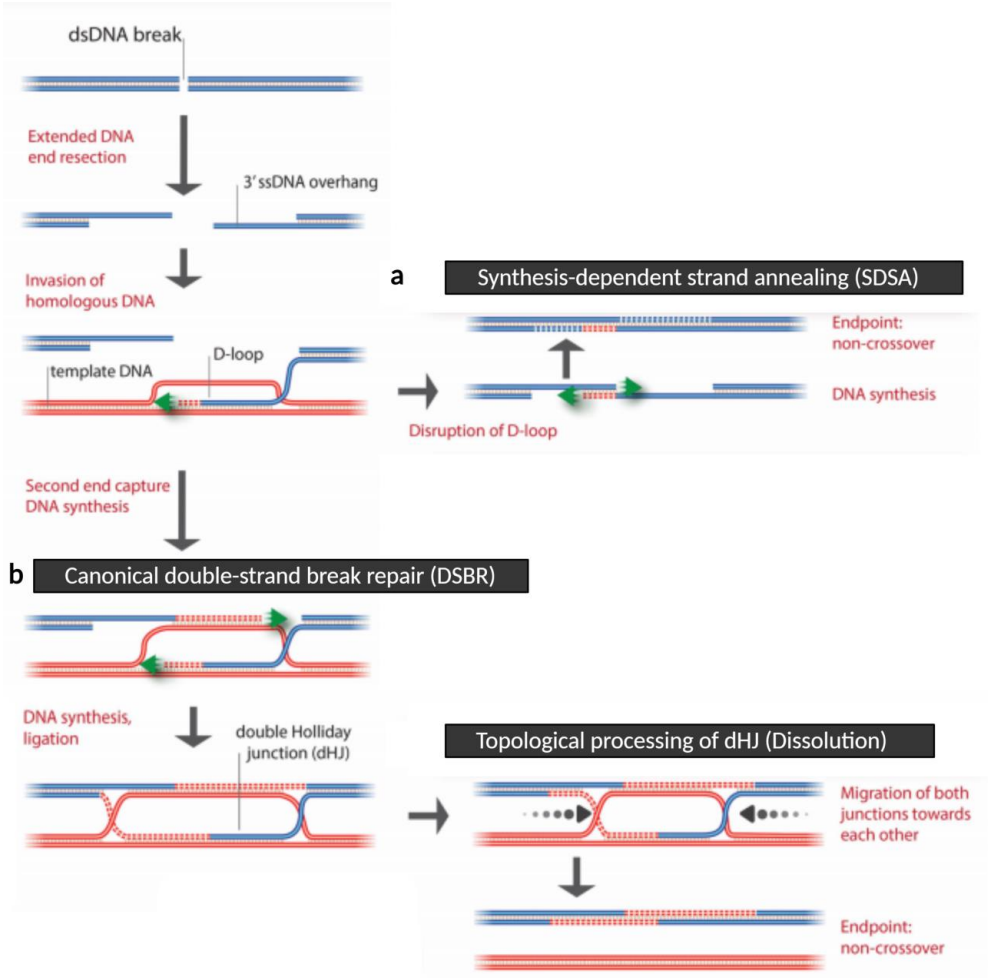


*Figure 1.22: General representation of the two HDR subpathways: SDSA (a) and DSBR (b). a) After strand invasion and DNA synthesis, if the D-loop is not stable, the broken DNA will anneal with the 3'ssDNA overhang at the other side of the break, fill the gaps and ligate the strands together. b) If the D-loop is stable, a second end capture occurs with the other 3'ssDNA overhang pairing with the displaced strand from the D-loop. This leads to formation of double Holliday junctions. After DNA synthesis and ligation, the junctions will be removed by dissolution in somatic cells. (**Adapted from Ranjha. 2018;** (121)**).***

Figure 1.22 - General representation of the two HDR subpathways: SDSA (a) and DSBR (b)

1.9.2d - Double Holliday junctions' dissolution

There are two ways to remove the dHJs: dissolution (149) or resolution (150). Dissolution separates the two DNAs without exchanging the flanking sequences (151), while resolution can lead to crossovers (150). In somatic cells, dissolution of dHJs is the default way of removing the junctions, therefore, it will be the only pathway discussed in this introduction. The main component necessary to remove the dHJs through dissolution, is a heterocomplex named the dissolvasome. The dissolvasome is composed of the BTRR complex (BML helicase, the topoisomerase IIIα and the RecQ-mediated instability factors RMI1 and RMI2; (149)). The helicase's role is to displace the two junctions towards each other, while the topoisomerase will create a nick in one DNA strand, allowing the other strand to pass through. Those two mechanisms together lead to DNA separation of the repaired strand from the template duplex (Figure 1.22b and Figure 1.23 (121, 149)).

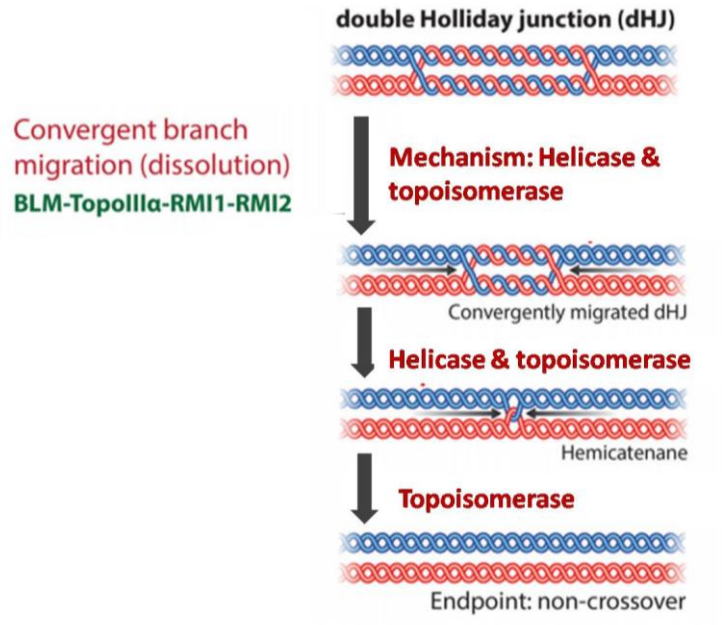


Figure 1.23: Graphic representation of the double Holliday junctions' dissolution. First, the BTRR complex enables the junctions to converge toward each other. Second, the topoisomerase IIIα cuts one of the strands, allowing separation of the two DNA duplexes. ***(Adapted from Ranjha et al. 2018;*** *(121)****).***

Figure 1.23 - Double Holliday junctions' dissolution

Because of the cell-cycle dependency, HDR efficiency is usually low during genome editing experiments in non-dividing human cells. Thinking in a therapeutic prospective for CF, the cell targets would most likely be airway epithelial cells, which contain both dividing cells (basal cells) and non-dividing cells. HDR could potentially be efficient in basal cells, with a group in 2020, showing HDR efficiency of about 20 to 40% in upper airway basal stem cells (152). However, this technique most likely will lead to potential *ex vivo* treatments, which is not well established nor particularly feasible at the moment (153).

If genome editing is to be used *in vivo*, using HDR pathway could be challenging as lung cells are mostly non-dividing cells. The use of HDR would lead to low correction efficiencies, since HDR is not active outside S and G2 phases.

1.10 HOMOLOGY-INDEPENDENT TARGETED INTEGRATION (HITI)

One technique to potentially solve this challenge was described in 2016 and uses the NHEJ pathway to perform precise genome editing with Cas9. This technique, called **H**omology-**I**ndependent **T**argeted **I**ntegration (HITI), is based on the design, into the donor plasmid, of a gRNA recognition site that is in the reverse-complementary sequence of the same recognition site in the genome (154). Usually, when NHEJ pathway is used to integrate a donor inside a cut genome, the donor has a 50% chance to integrate in the correct orientation, making the overall efficiency decreased by half. With the HITI design, when CRISPR/Cas system cut the sites both in the genome and in the donor plasmid, if the plasmid integrates inside the genome in the reverse orientation, two new gRNA recognition sites will be created on each side of the donor, making the Cas9 nuclease able to cut out the donor, and allowing it to integrate in the correct orientation (Figure 1.24).

There could be two designs for the donor: 1) the plasmid contains only one cut site, 2) the plasmid contains two cut sites flanking the sequence to be inserted. With the first plasmid, the insertion will lead to integration of both the sequence of interest and the plasmid. The advantage is that 100% of the integrated sequences will contain the sequence of interest. However, the disadvantage is that 100% of the integrated sequence will also contain the bacterial plasmid, which could lead to undesirable effects.

Using the second plasmid donor, the disadvantage is that there could be either integration of the sequence of interest, or integration of the plasmid, decreasing the editing efficiency. However, the advantage is that all the integrations of the sequence of interest will be plasmid free. Theoretically, with this HITI technique, the donor in the reverse orientation will be repeatedly excised until it inserts in the correct orientation. Consequently, the efficiency of integration is expected to be increased compared with a common NHEJ integration (154).

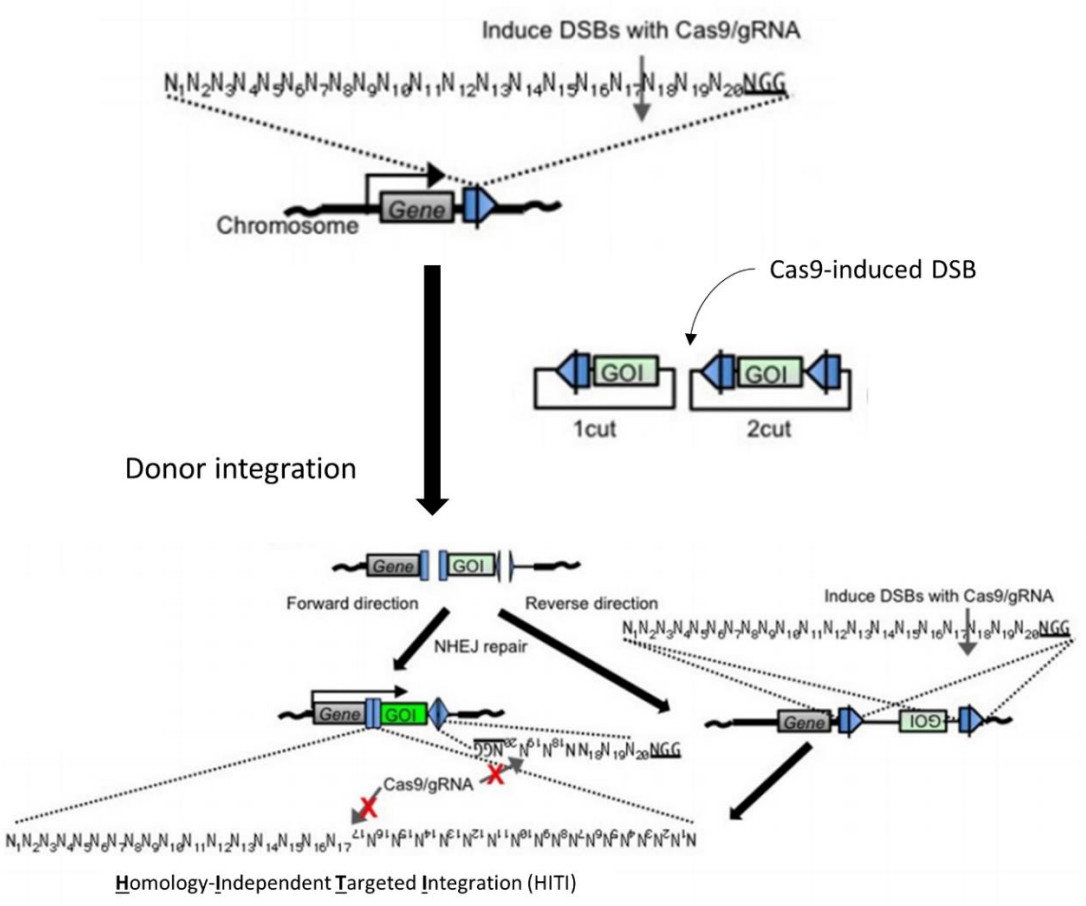


Figure 1.24: Graphic representation of the HITI principle. A donor plasmid was designed to contain the same guide recognition sites as the one used in the genome, but in the reverse-complementary orientation. There are two different donor designs, the donor could have only one gRNA recognition site, or two recognition sites flanking the sequence of interest. Upon Cas9 transfection, the nuclease cut both the genome and the donor plasmid. Because of the genomic double stranded break, the linearised plasmid can integrate in two possible orientations. On the left, the sequence of interest integrated in the correct orientation, disrupting the gRNA recognition sites and preventing Cas9 from cutting again. Thus, the integration is permanent and results in HITI. On the right, the donor integrated in the reverse orientation, forming two new gRNA recognition sites, which Cas9 can cut. The donor sequence is excised and has another chance to integrate in the correct orientation. ***(Figure adapted from Suzuki et al. 2016;*** *(154))*.

Figure 1.24 - HITI principle

At the moment the techniques for precise gene editing that were described here (HDR and HITI) all have the same limitations. Both of those techniques can lead to large indels in the genome. Indeed, if none of the donor are copied, for HDR, or integrated, for HITI, the DSB made by Cas9 will be repaired by NHEJ, creating indels. Those indels can damage the *CFTR* gene. One technique called Base Editing, designed in 2016, allows to overcome the problem of indels formations (section 1.11).

1.11 BASE EDITING

As seen previously, CRISPR/Cas9 system is a powerful tool to do genome editing. However, the off-target effects and indels formation could be a problem for its use in clinic. In 2016, a modified version of CRISPR/Cas9 technique, allowed conversion of one genomic base pair into another, at a specific genomic locus (155, 156). The technique, called base editing, does not need an exogenous DNA donor template and does not create double-stranded breaks, which are responsible for the potential harmful side-effects of genome editing such as big deletions or chromosomal rearrangements (157).

Base editing consists of the fusion of a deaminase on a catalytically inactive Cas9 protein, which has its nuclease domains deactivated to prevent DSB (**d**eactivated Cas9 (dCas9)). The dCas9 is still guided to the locus of interest by a gRNA, and the deaminase will convert cytidines into thymines (C>T), in the case of **C**ytidine **B**ase **E**diting (CBE), or adenines into guanines (A>G), in the case of **A**denine **B**ase **E**diting (ABE). The deamination occurs only in a specific region called “editing window”, corresponding to the positions relative to the protospacer that are susceptible to be converted. Base editors deaminate most or all nucleotides within the editing window.

1.11.1 - BASE EDITING FOR CYSTIC FIBROSIS

In the case of Cystic Fibrosis, around 66% of CF-causing mutations are point mutations. Amongst those point mutations, 15% can potentially be corrected by CBEs, and 46% by ABEs (Figure 1.25A). Amongst the latter are some class I nonsense mutations, which cause **P**remature **T**ermination **C**odon (PTC), leading to truncated messenger RNA and no or few functional proteins. There are no treatments yet available for those mutations (see section 1.3.3).

Using ABE technique, Lee et al. described in 2019, a technique called CRISPR-Pass, a method allowing to bypass those PTC by converting adenines (A), in the STOP codon, into guanines (G) (158). Indeed, the three stop codons are 5'-TGA-3', 5'-TAA-3' and 5'-TAG-3'. Using ABE, which converts A to G, the results of this conversion would lead to 5'-TGG-3' for all codons, which encodes a tryptophan (Trp) amino acid. Moreover, if ABE is used on the non-coding strand, the complementary sequences 3'-ACT-5', 3'-ATT-5' and 3'-ATC-5' codons, can be converted into 3'-GCT-5', 3'-GTT-5' and 3'-GTC-5' respectively. On the coding strand those conversions will give 5'-CGA-3', which encodes an Arginine (Arg), or 5'-CAA-3' and 5'-CAG-3', both encoding a Glutamine (Gln) (Figure 1.25B; (158)). Importantly, the conversion of the STOP codon to a Trp, Arg or Gln, will restore the original open reading frame of the gene and allow translation of a full protein.

However, if this new amino acid created is different from the original one in a WT healthy protein, studies will need to be done to confirm that the changed amino acid leads to a functional protein with no harmful side-effects.

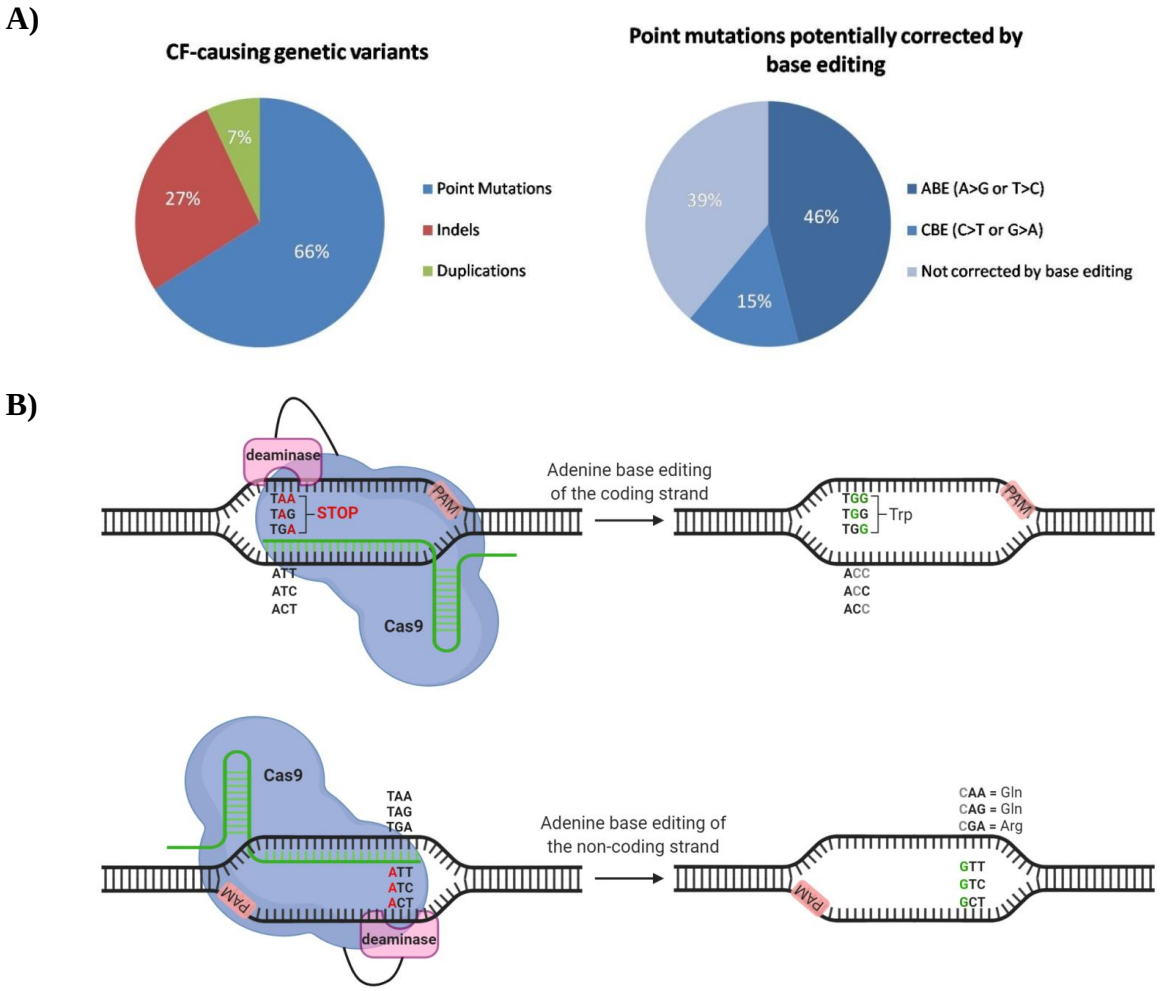


Figure 1.25: A) Graphic representation showing, on the left, the percentage of each type of mutation causing CF, and on the right, the percentage of point mutations that can be corrected by either CBE or ABE. ***(Figure from Mention et al. 2019;*** *(159)**).*** *B) Graphical representation of the CRISPR- Pass tool to bypass PTC. On the top is shown adenine base editing on the coding strand, while the bottom figure shows adenine base editing on the non-coding strand. Base editing on the coding strand lead to a conversion of STOP codons into tryptophan, while base editing on the non-coding strand can lead to a conversion into glutamine or arginine.*

Figure 1.25 - Percentages of point mutations that can be corrected by base editing and principle of the CRISPR- Pass tool to bypass PTC

1.11.2 - ADENINE BASE EDITORS (ABE)

Adenine Base Editors (ABEs) consist on the fusion between dCas9 and an adenine deaminase. The deaminase commonly used for ABE is a tRNA-specific adenosine deaminase (TadA) from *E. coli*, which was evolved in laboratory to recognise single-stranded DNA substrates and edit them with a high efficiency (156).

The dCas9:gRNA RNP will recognise the locus of interest and unwind the target DNA, creating a single-strand substrate for the deaminase, on the non-target strand. The deaminase will then recognise the adenines present in the editing window and convert them into inosines. Inosine (I) pairs with cytidine in the polymerase and is therefore read as a guanine (160). For the most common adenine base editors (ABE7.10 and ABEmax), the editing window corresponds to the nucleotides in position 4 to 7 in the protospacer, the PAM position being 21 to 23 (Figure 1.27; (156)). ABEmax is an evolved version of ABE7.10, in which the SV40 Nuclear Localisation Signal (NLS) has been replaced by a codon-optimised bis-bpNLS, therefore improving the nuclear localisation and increasing the editing efficiency from 1.3- to 7.9-fold (161).

1.11.3 - MISMATCH REPAIR (MMR)

Upon base conversion, a mismatch I:T is formed, which needs to be repaired by the cell. A mismatch in the genomic DNA can be repaired by two main pathways: 1) the post-replicative **M**is**M**atch**R**epair (MMR), or 2) the **B**ase **E**xcision **R**epair (BER) (162).

After replication, the MMR pathway corrects mismatches by differentiating the template DNA, from the new nascent DNA. To recognise the nascent DNA, the system will search for strand discontinuities in the replication fork and use DNA primers ends or Okazaki fragments as substrates for resynthesis (162). However, in base editing, there are no strand discontinuities, therefore the BER pathway is the major repair pathway. The BER pathway excises one of the bases and replace it with the correct pairing. However, because the system has no way to differentiate the faulty nucleotide, there will be a 50% chance for the edited base to be the one excised (162). Consequently, the efficiency of base editing is decreased. To improve the base editing efficiency, the group of David Liu, in 2016, proposed to create a nick in the target strand, creating a strand discontinuity, which will recruit the MMR pathway (155). To do so, a nickase Cas9 (nCas9) was used.

In the section 1.8.1, it was described that the Cas9 protein possesses two nuclease domains, RuvC and HNH. The HNH domain cleaves the target strand binding to the gRNA, and the RuvC domain cleaves the non-target strand. Mutations in one or the other domain lead to a nickase Cas9 (nCas9) which cleaves only one strand of DNA. The mutation D10A (nCas9 D10A) leads to the inactivation of the RuvC domain, making nCas9 D10A cutting only the target strand through the HNH domain. The mutation H840A (nCas9 H840A) inactivates the HNH domain, leading to the cleavage of the non-target strand through the RuvC domain (Figure 1.26; (163)).

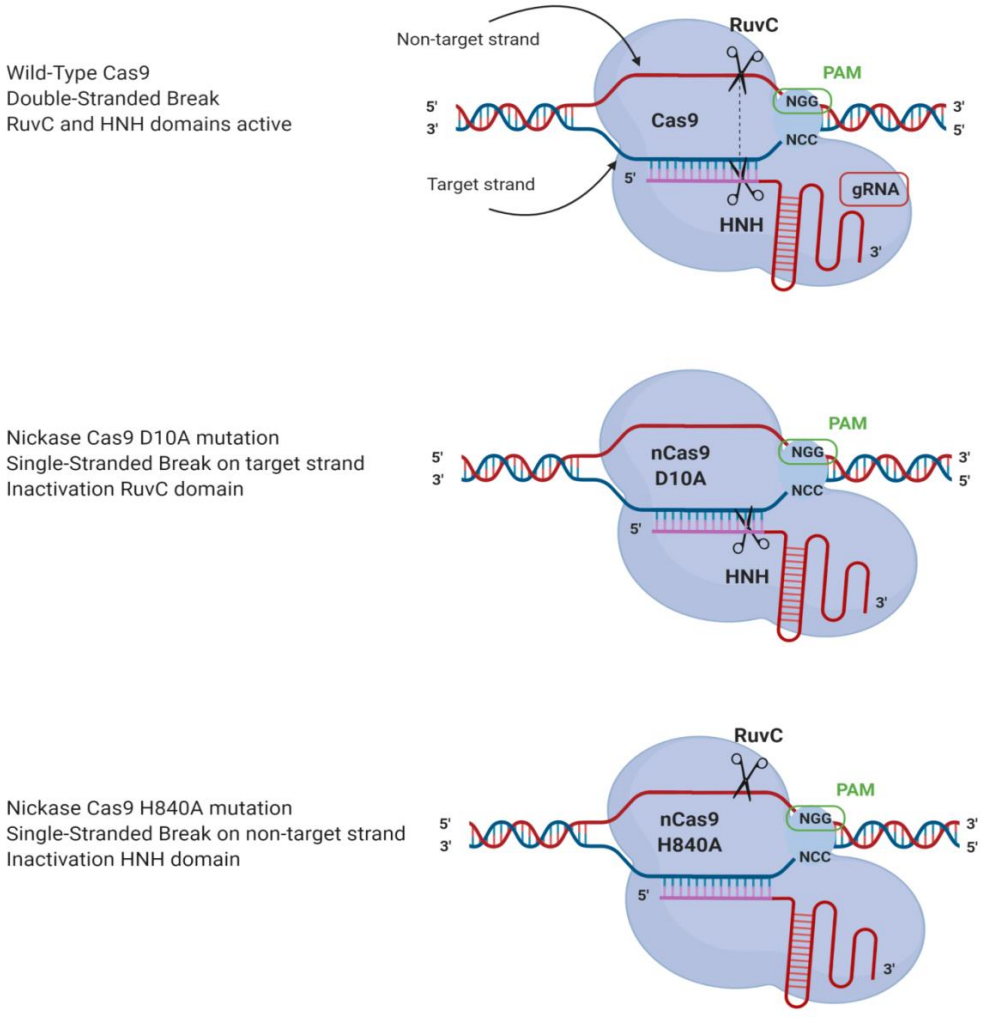


Figure 1.26: Schematic representation of SpCas9 nickases. The top picture corresponds to the WT SpCas9, which cuts both strands via the RuvC and HNH domains. The middle picture corresponds to the nCas9 D10A, which cuts only the target strand through the HNH domain. The bottom picture corresponds to the nCas9 H840A, which cuts only the non-target strand through the RuvC domain.

Figure 1.26 - SpCas9 nickases

To take advantage of the MMR pathway, the adenine deaminase was fused to the nCas9 D10A. The nick in the target strand, recruits the MMR factors, which will perform an end resection of the strand containing the mismatched thymine, and resynthesise a new DNA strand with the correct pairing (162). Therefore, the converted guanine will be conserved, and the MMR pathway will correct the base pairing. Using this technique, the efficiency of editing was improved (Figure 1.27; (155)).

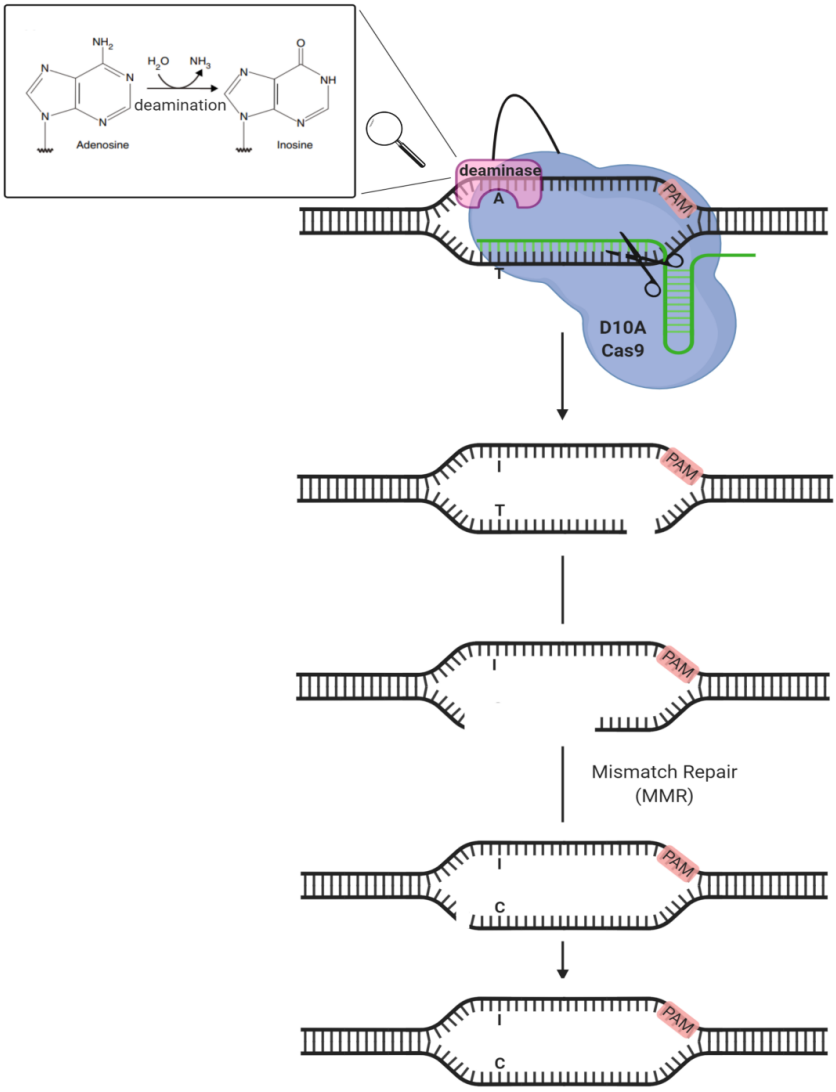


*Figure 1.27: Graphic representation of adenine base editing. The nCas9 D10A recognises the target locus, unwinds the DNA and cuts the target strand. The deaminase converts adenines into inosines in the 4 to 7nt editing window (**deamination image on the top left taken from Slotkin and Nishikura, 2013;** (164)). The MMR system recognises the cut strand, sections it, and resynthesises the complementary strand. Because the RNA polymerase recognises inosines as guanines, the original thymine is changed into a cytidine.*

Figure 1.27 - Principle of adenine base editing

1.11.4 - ADVANTAGES OF BASE EDITING

There are two main advantages for the use of base editing:

1) The efficiency of edition is high. For example, in 2018, Koblan et al. used the optimised adenine base editor ABEmax, and demonstrated about 30% to 46% base editing of the *HGB1/2* gene, encoding γ-globin, without pre-selection of the cells (161). Another example, with the cytidine base editor, in 2016, Komor et al. compared BE3 with a conventional HDR technique, using SpCas9 and ssDNA as a donor template, to introduce the same base modification. Comparing two genomic loci *APOE4* and *TP53*, they found 50% and 7.6% editing respectively, using BE3, and 0.3% and non-detectable editing respectively, using HDR (Table 1.2; (155)). This study showed that base editing is much more efficient than the conventional HDR technique.

2) Because there is no DSB, base editing is safer to use. Indeed, it has been shown that the DSB created by Cas9 can lead to big deletions and chromosomal rearrangements (157). Moreover, base editing leads to very low levels of indels compared to canonical SpCas9 HDR editing. The same study on *APOE4* and *TP53* demonstrated for HDR up to 40% and 8% indels respectively, while using base editors, up to 6% and less than 0.7% indels for each gene respectively were observed (Table 1.2; (155)). Concerning adenine base editors the ABE7.10 showed <0.1% indels (156), while the most efficient version ABEmax displayed <1.6% indels (161). Despite a clear increase in indels with ABEmax compared to ABE7.10, the amount of indels is still very low compared to conventional SpCas9 HDR. Because indels can lead to frameshift modifications and knock-out of a protein, having a technique which creates very low indels would be safer for a potential use in the clinic.

	APOE4		***TP53***	
	Editing	**indels**	**editing**	**indels**
Cas9 + ssDNA donor (HDR)	0.1-0.3%	26-40%	ND	6.1-8%
BE3	<50%	4.6-6.1%	3.3-7.6%	<0.7%

Table 1.2: Table showing percentages of editing efficiency and indels formation, for two different genes (APOE4 and TP53), using either the cytidine base editor BE3 or the conventional SpCas9 HDR with ssDNA donor template (155).

Table 1.2 - Percentages of editing efficiency and indels formation, for two different genes (APOE4 and TP53), using either the cytidine base editor BE3 or the conventional SpCas9 HDR with ssDNA donor template

1.11.5 - LIMITATIONS OF BASE EDITING

There are four main limitations for the use of base editing:

1) Base editing is a very precise technique for genome editing. Indeed, it targets adenines in a short editing window and edits most to all the bases inside the window. However, most of the time, only one base needs to be changed, and if the editing window contains many adenines (for ABE), there could be conversions of unwanted bases. In figure 1.28 is shown an example of unwanted base editing. In A), the graph shows that the window of editing for the ABE 7.10 is from position 4 to 7 (Figure 1.28A). In B), Ryu et al. (165) used a protospacer with many adenines inside. The group showed the frequency of base conversions they observed. The most common conversion was a 38.75% A>G conversion at position 5. However, the second most common conversion was 22.59% of both A>G conversions at position 5 and 7 (Figure 1.28B). As shown here, the frequencies of both conversions are significantly high. Therefore, researchers will need to be careful with the unwanted editing and studies will need to be performed to confirm the safety of those editing.

2) Two base editors are currently available. Cytidine base editors can convert C>T and G>A, while adenine base editors can convert A>G and T>C. Therefore, there are still some conversions that are not available. Consequently, there could be many cases were the editing will lead to a change in amino acid sequence compared to the WT protein. If the amino acid needs to be changed, analyses will need to be performed to check for functionality and safety.

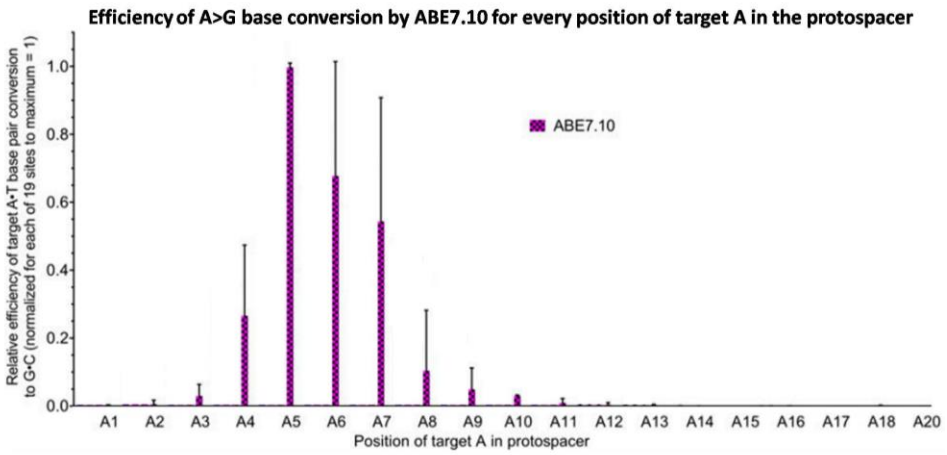
3) PAM limitation, as discussed previously for Cas9, is also a problem for base editing. Base editing having a short editing window of 4 nucleotides, it is important to have the PAM at positions 14 to 17nt away from the base that needs to be converted. To overcome the limitation of the PAM, a group fused the most efficient adenine base editor ABEmax (161), with an engineered SpCas9-NG protein (166) that can recognise 5'-NG-3' PAM (167). With this new base editor, the site limitation should be significantly decreased.

4) A recent study showed that both CBEs and ABEs creates RNA off-targets (168). Some RNAseq studies showed that the use of BE3 and ABEmax altered the sequence of many RNAs, showing more than 10,000 RNA base positions edited from A to G. The RNA edits concerned both coding and non-coding sequences. More than 50% of RNA gene transcripts contained at least one edited adenine, and 95% of those RNA editing were not present in the DNA (168).

Researchers are advised to use a transient base editor that could edit RNA in a very short period of time, to restrict damages. Also, transcriptomic analyses will need to be done to assess the consequences of the RNA edits. Finally, a **S**elective **C**urbing of **U**nwanted **R**NA **E**diting (SECURE) APOBEC-based cytidine base editor has been designed to reduce those RNA edits made by cytidine deamination (168). Two variants of the BE3 (BE3-R33A and BE3-R33A/K34A) displayed comparable on-target efficiencies than the BE3, however, the editing was more precise, with the first variants showing only hundreds of off-target RNA editing, and the second variant showing 26 or lower (168).

Few months later, the SECURE-ABE was designed to reduce RNA off-targets produced by ABE (169). To create SECURE-ABE, Grünewald et al. truncated ABEmax by removing the WT *E.Coli* TadA domain that was believed to retain its original function of RNA deaminase. Because this alone was not sufficient to reduce the off-target RNA editing, two variants were engineered from this truncated ABEmax; K20A/V82G and R21A/V82G (169). Those variants showed a slight decrease in the on-target editing compared to ABEmax; however, the group also noticed a substantial reduction in off-target RNA editing. Unfortunately, the off-target RNA editing was still about 3 to 4 times higher than the background, demonstrated that more work still needs to be done to further decrease off-targets (169).

A)



B)

AAGGAAACTG**GAACACAAAGCATAGACTGC**GGG	Frequency (%)
AAGGAAACTGGAACGCAAAGCATAGACTGCGGG	38.75
AAGGAAACTGGAACGCGAAGCATAGACTGCGGG	22.59
AAGGAAACTGGAACGCAGAGCATAGACTGCGGG	0.73
AAGGAAACTGGAACGCAAGGCATAGACTGCGGG	0.58
AAGGAAACTGGAACGCGGAGCATAGACTGCGGG	0.44

Figure 1.28. Graphic representation showing the limitations caused by the editing window, for ABE7.10. A) Graph showing the efficiency of conversion at each position in the protospacer, for ABE7.10. The highest editing efficiencies are for adenines in positions 4 to 7nt. ***(Figure adapted from Gaudelli et al. 2017*** *(156)****)****. B) Example of editing frequencies at the HEK2 site. The protospacer possesses many adenines. The highest frequency of correction is at position 5. However, a large frequency of conversions also happened at position 7.* ***(Figure adapted from Ryu et al. 2018*** *(165)****)****.*

Figure 1.28 - Base editing limitation caused by the editing window, for ABE7.10

1.11.6 - CONCLUSION ON BASE EDITING

Base editing is a powerful technique for high efficiency conversion of one base. However, for Cystic Fibrosis, there are more than 300 different mutations. About 30% of those mutations could be corrected with ABE. Four of the most common mutations (G551D (2.1%); N1303K (1.5%); R117H (1.3%) and W1282X (1.2%)) could theoretically be corrected by ABE and having their original sequence restored. Moreover, the second most common mutation G542X (2.5%), which is a nonsense mutation, can also have its open reading frame restored using ABE. However, base editing of this mutation will lead to an amino acid change, which will need to be analysed for safety. For those mutations, the use of ABE seems to be a safe and efficient method. However, for other mutations, which are rarer, designing base editors and new guides for each mutation could be money-consuming for companies which might want to use the money for more frequent mutations or diseases. Therefore, there is still a need for the development of techniques for rarer mutations.

1.12 SUPEREXON

In 2011, a clinically relevant genome editing experiment in a mouse model of Haemophilia B, using ZFN, has been published (170). In this publication, the authors used a technique now called “superexon”, where the partial cDNA of the human *F9* gene, from exon 2 to 8, was designed, flanked by a splicing acceptor site upstream and a poly-adenylation site downstream (Figure 1.29B). This superexon was inserted through HDR into the first intron of the *F9* gene (Figure 1.29A). Using this technique, the correctly inserted sequences were spliced to the endogenous exon 1, leading to a cDNA with the endogenous exon 1, and the exogenous exons 2 to 8. Using this technique, every mutation in the *F9* gene localised downstream of exon 1, should be corrected with this unique construct.

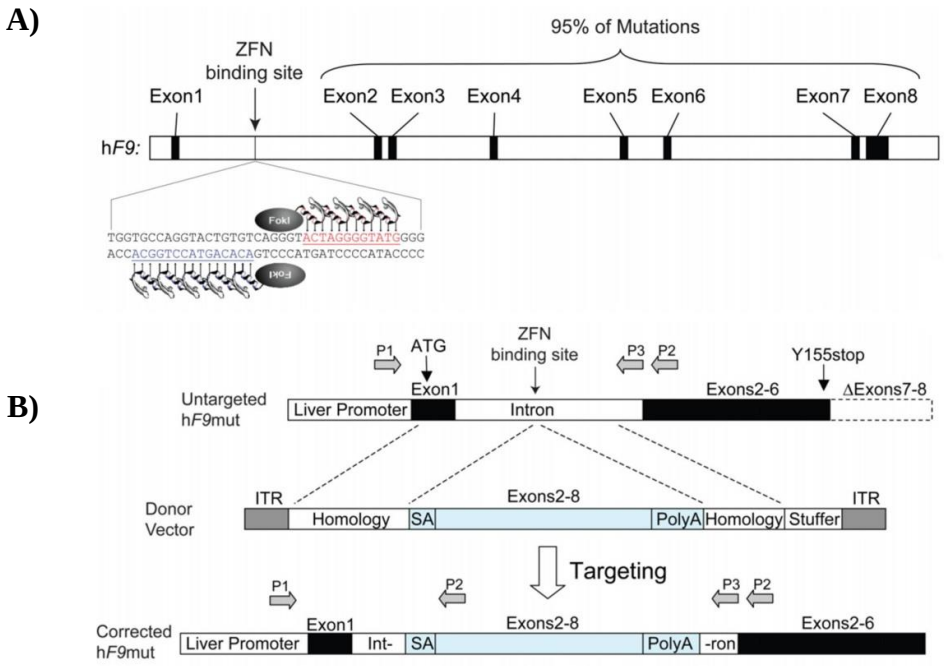


Figure 1.29: Gene correction of the hF9 gene using a superexon. A) graphic representation of the human F9 gene. The scheme shows the exons and introns, as well as the location of the ZFN binding sites. B) Graphic representation of the donor vector and the result of its integration into the mutated gene. ***(Adapted from Li et al. 2011*** *(170))*.

Figure 1.29 - Gene correction of the hF9 gene using a superexon

Using this superexon technique for Cystic Fibrosis would allow correction of the majority of the mutations, even the rarest. Due to the design of only one unique donor to correct any mutations, the superexon technique offers the opportunity to treat everybody.

In 2016, Bednarski et al. used the superexon technique to insert a superexon containing *CFTR* exons 11 to 27, inside the exon 11 of *CFTR,* in CFBE41o$^-$ cells (171). A ZFN pair was designed to cut the exon 11 of the CFTR gene. The donor was composed of 690bp of left homology arm homologous to the intron 10 and 840bp of right homology arm homologous to exon 11 and parts of intron 11. The sequence corresponding to the ZFN recognition site in the donor, contained silent mutations to prevent further cutting after integration, and facilitate detection. Between the homology arms was designed the superexon 11-27 followed by a puromycin selection gene driven by a PGK promoter (171). The group showed successful insertion of the superexon by HDR. Moreover, the superexon integration showed restoration of CFTR function in a F508del cell line. However, the integration efficiency using HDR was very low, with 10% integration after a puromycin selection (171).

AIMS

In this thesis project, the overall purpose is to compare different genome editing techniques developed and optimised during the recent years, to determine if there is one or several that could be optimal for correcting CF-causing mutations, using W1282X variant as a model system. The hypothesis is that at least one of those technique would correct the *CFTR* gene, restore mRNA stability and protein production and function.

The first two objectives will focus on correcting the single mutation W1282X using HDR with AsCas12a and Cas9 (chapter 3) or using adenine base editor (chapter 4).

In chapter 3, the W1282X mutation will be corrected with AsCas12a, using electroporation in 16HBE14o$^-$ cells. In chapter 4, the mutation will be corrected using the NG-ABEmax base editor, using Lipofectamine in HEK FLP W1282X EMG in21-in24 cells.

The final two objectives will focus on correcting every mutation localised in exon 23 and downstream, comprising W1282X, using a superexon.

In chapter 5, the homology-independent targeted integration (HITI) will be used to physically integrate a superexon comprising exon 23 to 27 into the exon 22 of the *CFTR* gene, using an existing gRNA cutting exon 22. The superexon plasmid as well as the Cas9/gRNA plasmid were transfected using Lipofectamine in both HEK293 and 16HBE14o$^-$ cells.

In chapter 6, HDR was used to replace the whole exon 23 (156bp) with a template containing a codon-optimised exon 23 with the W1282X correction, flanked by homology arms. The template plasmid and the Cas9/gRNA plasmid were transfected using Lipofectamine in both HEK293 and HEK FLP W1282X EMG in21-in24 cells.

The four techniques are based on different repair pathways such as HDR (chapter 3 and 6), MMR (chapter 4) and NHEJ (chapter 5).

In conclusion of this project, the advantages and limitations for each technique will be displayed to allow researchers to have a clear overview of gene editing for CF research in order to optimise them and eventually bring them into potential clinical trials.

CHAPTER 2 - MATERIALS AND METHODS

2.1 CELL LINES

HEK293: Human embryonic kidney 293 cell line (172). Hypotriploid human cell line. Easy to transfect both with liposomes and electroporation systems. HEK293 cells do not express *CFTR* mRNA or protein (173). Moreover, the cells do not have any ability to form a resistant monolayer necessary to measure ion transport. The cells were kindly given by Dr Scallan (University College Cork, Cork, Ireland)

16HBE14o$^-$: Human bronchial epithelial cells. The cells express high levels of *CFTR* mRNA and protein. They are able to form a differentiated epithelial monolayer with transepithelial resistance (174). The 16HBE14o$^-$ cells and their mutants were shown to possess an insertion of unknown size inside intron 6 of one of the *CFTR* allele. The insertion corresponded to a SV40 fragment used for the immortalisation process of the cells. As a result, an aberrant transcript is produced leading to a non-functional CFTR protein. Therefore, only one of the *CFTR* allele is functionally active (175). The cells were bought from the Cystic Fibrosis Foundation Lab (Lexington, MA, USA).

16HBE14o$^-$ W1282X mutant cells: 16HBE14o$^-$ cells genetically modified to be homozygous for the W1282X mutation. The cells possess the mutation c.3846G>A in exon 23 of the *CFTR* gene (175). The cells were bought from the Cystic Fibrosis Foundation Lab (Lexington, MA, USA).

HEK Flp-In 293 EMG i21-i24 W1282X: HEK 293 containing a **F**lp **R**ecombinase **T**arget (FRT) were transfected with the **E**xpression **M**ini**G**ene (EMG) pcDNA5/FRT/hygro/pCMV-CFTRi21-24 and the pOG44 recombinase plasmid. This system allows the Recombinase to integrate the EMG into the FRP site in HEK cells. The EMG i21-i24 has been designed by Aksit et al. and is composed of the *CFTR* cDNA with shorter introns 21 to 24. Intron 21 contains 227nt from the 5' end and 222nt from the 3'end, intron 22 contains 190nt from the 5' end and 257nt from the 3' end, intron 23 contains 314nt from the 5' end and 327nt from the 3' end, finally, intron 24 contains 339nt from the 5' end and 312nt from the 3' end (176). The HEK Flp-In EMG i21-24 W1282X cells possess a single active *CFTR* allele under the control of the CMV promoter, which can be expressed as a *CFTR* mRNA.

However, because they are HEK cells, the cells do not form a highly resistant polarised monolayer for Ussing Chamber functional assays. The cells were kindly given by Neeraj Sharma (Johns Hopkins University, Baltimore, MD, USA).

2.2 CELL CULTURE

2.2.1 - CELL CULTURE REAGENTS

Culture medium:

- Minimum Essential Medium Eagle (MEM, with Earle′s salts and sodium bicarbonate, without L-glutamine, liquid, sterile-filtered; Sigma-Aldrich), supplemented with 10% (v/v) Foetal Bovine Serum (FBS, non-USA origin, sterile-filtered; Sigma-Aldrich), 2mM L-Glutamine (200mM, solution, sterile-filtered; Sigma-Aldrich), 500 units of penicillin and 50mg of streptomycin (Pen-Strep, solution stabilised, with 100 units penicillin/ml and 10 mg streptomycin/ml, sterile-filtered; Sigma-Aldrich).

Freezing medium:

- Culture medium with 10% (v/v) dimethyl sulfoxide (DMSO, Sigma-Aldrich).

Other reagents:

- Trypsin-EDTA solution (1X, sterile-filtered, 0.5g porcine trypsin and 0.2g EDTA; 4Na/L of Hanks′ Balanced Salt Solution with phenol red; Sigma-Aldrich)
- 1X PBS was prepared from a dilution 1:10 of 10X PBS (0.1M Phosphate Buffer, 1.54M Sodium Chloride; Sigma-Aldrich) with autoclaved deionised water.
- Mycoplasma treatment detection primers (177) (Eurofins):

GPO-1 FW:	5‘-ACTCCTACGGGAGGCAGCAGTA-3‘
MGSO REV:	5‘-TGCACCATCTGTCACTCTGTTAACCTC-3‘

- Mycoplasma treatment solutions**:** Myco-1, Myco-2 and Myco-3 (AppliChem)

2.2.2 - PASSAGE OF THE CELLS

To passage the cells from a T75 flask, the medium was removed, and the cells were washed with 5ml of 1X Phosphate Buffered Saline (PBS) solution (for HEK cells), or 2ml of warm Trypsin-EDTA solution (for HBE cells). Following the wash, 2ml of Trypsin-EDTA solution were added to the flask and incubated 3 min at 37°C for HEK cells or 15min for HBE cells. To inactivate the trypsin, the culture medium was added to the cells to reach a final volume of 10ml. Nine millilitres of the mix were removed to keep $1/10^{th}$ of the cells and 9ml of fresh culture medium were added to the flask. Cells were incubated at 37°C with 5% (v/v) CO_2. The flask was changed after every 5 passages and the cells were discarded and replaced by new freshly thawed ones after 25 passages.

2.2.3 - CELL STORAGE

The cells were stored in liquid nitrogen. Once the flask reached about 90% confluency, the cells were trypsinised as in section 2.2.2 above. Upon trypsinisation, the cells were centrifuged at 1000rpm for 5min. The pellet was resuspended with 1ml freezing medium per tube to freeze. The cells were added to a Biosphere® SC Micro Tube 2.0ml (Sarstedt) and directly placed into a freezing container half filled with isopropanol and incubated at -80°C for up to 1 week, before placing them in the liquid nitrogen for longer conservation.

2.2.4 - MYCOPLASMA SCREENING AND TREATMENT

Every two months, the cells in cultures were screened for mycoplasma. Before assessment, the cells were cultured without antibiotics for at least one week. One millilitre of medium was collected and incubated at 95°C for 10min. The samples were centrifuged at 200g for 1min and 1µl of supernatant was used for PCR (sections 2.5.2 and 2.5.6).

PCR reaction per sample:	
5X Green GoTaq® Flexi Buffer	5µl
25mM $MgCl_2$	2.5µl
dNTP	1µl
Primer GPO Fw (10µM)	0.5µl
Primer MGSO Rev (10µM)	0.5µl
GoTaq® Flexi DNA Polymerase	0.25µl
Water	15.25µl
Supernatant	1µl

PCR Program:			
Initial denaturation	96°C	5 min	
Denaturation	94°C	30 sec	**40 Cycles**
Annealing	55°C	30 sec	
Extension	72°C	1 min	
Final extension	72°C	10 min	
Hold	4°C	∞	

The PCR product was analysed on an electrophoresis 1% (w/v) agarose gel (materials and methods in sections 2.5.2 and 2.5.4b).

Cells positive for mycoplasma were discarded if an uncontaminated stock was available. Otherwise, the cells were treated as follow:

One percent of a myco-3 solution was added to the culture medium and was renewed every 3 days for 2 weeks. After those 2 weeks, the cells were incubated for another two weeks with antibiotic-free and treatment-free medium. Another mycoplasma test was completed after treatment. If the cells were still positive, a supplementary treatment with myco-1 and myco-2 was done sequentially.

At first, 1% (v/v) of myco-1 was added to the culture medium for 4 days, then, 1% (v/v) of the myco-2 solution was added to the culture medium for another 3 days.

2.3 CELL TRANSFECTION

2.3.1 - REAGENTS

Transfection media:

- Transfection medium 1 (TFX1): Minimum Essential Medium Eagle (MEM, with Earle′s salts and sodium bicarbonate, without L-glutamine, liquid, sterile-filtered; Sigma-Aldrich), supplemented with 10% (v/v) Foetal Bovine Serum (FBS, non-USA origin, sterile-filtered; Sigma-Aldrich) and 2mM L-Glutamine (200 mM, solution, sterile-filtered; Sigma-Aldrich).
- Transfection medium 2 (TFX2): Minimum Essential Medium Eagle (MEM, with Earle′s salts and sodium bicarbonate, without L-glutamine, liquid, sterile-filtered; Sigma-Aldrich), supplemented with 2mM L-Glutamine (200 mM, solution, sterile-filtered; Sigma-Aldrich).
- Opti-MEM®: Reduced Serum Medium (Gibco™, ThermoFisher)

Lipofectamine:

- Lipofectamine 2000 Transfection Reagent (Invitrogen)
- Lipofectamine 3000 Transfection Reagent (Invitrogen)
- Lipofectamine LTX Reagent with PLUS Reagent (Invitrogen)

Electroporation:

- Neon Transfection System (ThermoFisher Scientific)
- Neon™ Transfection System 10 µL Kit (ThermoFisher Scientific). Contains Resuspension Buffer R, Electrolytic Buffer E, ninety-six 10µl Neon® tips and 20 Neon® electroporation tubes.
- Dulbecco's Phosphate Buffered Saline (DPBS, modified, without calcium chloride and magnesium chloride, liquid, sterile-filtered; Sigma-Aldrich).

2.3.2 - TRANSFECTION USING LIPOFECTAMINE

For every Lipofectamine transfection, cells were plated the day before. HEK293 cells were plated at a number of 10^5 cells in a 24-well plate or $2x10^5$ cells in a 12-well plate. 16HBE14o$^-$ cells were plated at a number of $3x10^5$ cells in a 24-well plate or $5x10^5$ in a 12-well plate. During the incubation time of the samples, the cell medium was changed to TFX1 medium.

2.3.2a - Lipofectamine 2000

For a 12-well plate, 4µl of Lipofectamine 2000 were mixed into 100µl of TFX2 medium. The Lipofectamine mix was incubated at room temperature (RT) for 5min. One hundred microlitres of this Lipofectamine mix were added to 1µg to 1.5µg of DNA diluted in up to 100µl of TFX2. After 20min incubation, 200µl of mix were slowly added to the cells.

2.3.2b - Lipofectamine 3000

For a 12-well plate, 2µl of Lipofectamine 3000 in 50µl of TFX2 were added to a mix containing 1µg to 1.5µg of DNA and 2µl of P3000 reagent, diluted in up to 50µl of TFX2. After 15min incubation, 100µl of mix were slowly added to the cells.

2.3.2c - Lipofectamine LTX

For a 24-well plate, 3µl of Lipofectamine LTX in 25µl of TFX2 were added to a mix containing 0.8µg of DNA and 2.4µl of Plus reagent, diluted in up to 25µl of TFX2. After 5min incubation, 50µl of mix were slowly added to the cells.

2.3.3 - TRANSFECTION USING ELECTROPORATION

The plate receiving the cells was filled with 1ml of TFX1 in each well of a 12-well plate (500µl for 24-well plate and 2ml for 6-well plate) and put in the incubator, at 37°C, until the electroporation. Cells were trypsinised like in section 2.2.2. After trypsinisation, cells were centrifuged for 5min at 1000g, washed with 5ml of DPBS and centrifuged again. The cells were, then, resuspended with enough Buffer R to have the right number of cells for each condition.

2.3.3a - Electroporation of plasmids

In a 12-well plate, $4x10^5$ HBE cells were electroporated with 3.25µg of plasmid DNA at 1350V, Width 10ms and 3 Pulses.

In a 12-well plate, $2.5x10^5$ HEK cells were electroporated with 1µg of plasmid DNA at 1150V, Width 20ms and 2 Pulses.

2.3.3b - Electroporation of Ribonucleoproteins (RNP)

2.3.3.b-i Electroporation of Cas9 RNP

The electroporations were done following the IDT protocol: *Alt-R CRISPR-Cas9 System: Delivery of ribonucleoprotein complexes into Jurkat T cells using the Neon® Transfection System (IDT).*

Alt-R® S.p. HiFi Cas9 Nuclease V3 (IDT) was diluted to a 36µM concentration. To form the RNP complex, the same volume of crRNA:tracrRNA annealed guides (40µM) and Cas9 nuclease was added to have a final concentration of 18pmol of Cas9 nucleases and 20pmol of guides.

In a 12-well plate, $2.5x10^5$ **HEK** cells were electroporated with 1µg of plasmid DNA, 1µl of Cas9-RNP mix (final concentration 1.6µM gRNA and 1.5 Cas9) at 1150V, Width 20ms and 2 Pulses.

In a 12-well plate, $2.5x10^5$ **HBE** cells were electroporated with 1µl of Cas9-RNP mix (final concentration 1.6µM gRNA and 1.5µM Cas9) at 1290V, Width 20ms and 2 Pulses. If a donor plasmid was used with the RNP electroporation, different concentrations of plasmids were added after the RNP and the cells, just before the electroporation.

2.3.3.b-ii Electroporation of Cas12a RNP

The electroporations were done following the IDT protocol: *Alt-R® CRISPR-Cas12a (Cpf1) System: Delivery of ribonucleoprotein complexes into Jurkat T cells using the Neon® Transfection System (IDT).*

To form the RNP complex, the same volume of Alt-R® A.s.Cas12a (Cpf1) Nuclease V3 (IDT) and crRNA (75µM) was added to have a final concentration of 75pmol of Cas12a and 63pmol of crRNA. In a 24-well plate, $2x10^5$ HBE cells were electroporated with 2µl of Cas9-RNP mix (final concentration 6.25µM gRNA and 5.25µM Cas9) at 1290V, Width 20ms and 2 Pulses. If a donor plasmid was used with the RNP electroporation, different concentrations of plasmids were added after the RNP and the cells, just before the electroporation.

2.4 MICROBIOLOGY

2.4.1 - BACTERIAL STRAINS

- 5-alpha Competent *E. coli* (High Efficiency) DH5α (New England BioLabs, NEB)

2.4.2 - MICROBIOLOGY REAGENTS

Growth Media:

- Lennox Broth (LB) 100 tablets (Sigma-Aldrich): one tablet per 50ml deionised water was autoclaved.
- Lennox Broth with Agar (LB) 100 tablets (Sigma-Aldrich): one tablet per 50ml deionised water was autoclaved. After the bottle cooled down, the antibiotic was added, and the medium was poured into Petri dishes to solidify.

Other Reagents:

- Glycerol 50% (v/v): 200ml glycerol (Sigma-Aldrich) mixed with 200ml of autoclaved deionised water.
- Ampicillin sodium salt (Sigma-Aldrich): a 50mg/ml stock solution was made in deionised water. The solution was filtered, and aliquots were kept at -20°C.
- Kanamycin sulfate from *Streptomyces kanamyceticus* (Sigma-Aldrich): a 50mg/ml stock solution was made in deionised water. The solution was filtered, and aliquots were kept at -20°C.

2.4.3 - BACTERIA TRANSFORMATION

Fifty microlitres of competent bacteria were mixed with 5µl of plasmid DNA. After 30min of incubation on ice, a heat-shock was performed by putting the cells at 42°C in a heat block for 40sec and put on ice again for 5min. For DH5α *E. coli*, 400µl of LB medium was added to the tube, and 50µl were spread on a pre-warmed agar plate containing the relevant antibiotic. The plates were incubated at 37°C overnight.

2.4.4 - MINIPREP

All minipreps were performed using the NucleoSpin® Plasmid DNA Purification Kit (Macherey-Nagel), and following the protocol from the kit. From the bacteria transformation (see section 2.4.3), several colonies were put individually into culture in 5ml LB medium with the relevant antibiotic. The cultures were shaken at 250rpm, overnight at 37°C. From those 5ml cultures, 1.5ml were taken and placed in an Eppendorf tube. Tubes were centrifuged at 11000g for 30sec, to pellet bacteria. The pellets were resuspended with 250µl of buffer A1, before the bacteria were lysed for 5min using the buffer A2. The lysis was stopped using the neutralising buffer A3, and the debris were pelleted by centrifugation 5min to 10min at 11000g. The supernatants were added to NucleoSpin columns and centrifuged for 1min at 11000g. The columns were washed using 600µl of buffer A4 and were centrifuged for 1min at 11000g. A second centrifugation was done for 2min, to dry the membrane and remove any traces of ethanol. The plasmids' elutions were performed using 50µl of buffer AE and centrifuged for 1min at 11000g.

2.4.5 - MIDIPREP

2.4.5a - Midiprep QIAGEN® plasmid midi kit

Midipreps were performed using the QIAGEN® Plasmid Plus Midi Kit (Qiagen) and following the protocol from the kit. From the 5ml cultures (section 2.4.4), 50µl were put into 50ml of LB medium with the proper relevant antibiotic, in a 250ml Erlenmeyer. The culture was shaken overnight at 250rpm, 37°C. The 50ml were centrifuged at 4500g for 20min, 4°C, and the pellet was resuspended with 4ml of buffer P1. The cells were then lysed with 4ml of buffer P2 for 3min, and the lysis was stopped using 4ml of the neutralising buffer S3. The lysate was transferred into a QIAfilter Cartridge, and incubated for 10min at RT. Using a plunger, the lysate was filtered through the cartridge, and 2ml of buffer BB were added. The mixture was added to a QIAGEN Plasmid Plus spin column fixed to a QIAvac 24 Plus. Three hundred millibar vacuum was applied to suck the liquid out the column. Seven hundred microlitres of buffer ETR (endotoxin wash buffer) were added, followed by 700µl of Buffer PE to wash the DNA. The column was removed from the QIAvac and centrifuged for 1min at 10,000g to remove any trace of ethanol.

The plasmid was then eluted with 200µl of buffer EB, by centrifugation for 1min at 10,000g. The Qiagen midiprep kit, enables reduction of endotoxins, however, it does not lead to a completely endotoxin-free preparation.

2.4.5b - MIDIPREP NUCLEOBOND® XTRA MIDI EF

Midipreps were done using the endotoxin-free plasmid DNA purification kit (NucleoBond® Xtra Midi EF; Macherey-Nagel) and following the protocol from the kit. After pelleting the 50ml culture as in section 2.4.5a, the pellet was resuspended using 8ml of buffer RES-EF and lysed using 8ml of buffer LYS-EF for 5min. The lysis was neutralised using 8ml of buffer NEU-EF, and the lysate was incubated on ice for 5min. The lysate was poured into the NucleoBond® Xtra Column filter and filtered by gravity. The first wash was done using 5ml of buffer FIL-EF. The column was discarded, and a second wash of the filter was done using 35ml of buffer ENDO-EF. A last wash was done using 15ml of buffer Wash-EF, before elution with 5ml of buffer ELU-EF. The DNA was measured using a NanoDrop® Spectrophotometer ND-1000 (ThermoFisher). Three point five millilitres of isopropanol were added to the eluate, vortexed and incubated at RT for 2min before being loaded to a NucleoBond® Finalizer. The flow-through was discarded and the filter was washed with 2ml of 70% (v/v) ethanol and dried by forcing the air through the filter.

Depending on the amount of DNA measured by the nanodrop, a certain amount of elution buffer TE-EF was loaded, and the eluate was poured into an individual endotoxin-free Eppendorf tube. The amount of TE-EF buffer was determined to get around 1µg/ml of plasmid DNA in the eluate, following the kit table (Table 2.1).

Loaded DNA	Elution volume					
	100 µL	200 µL	400 µL	600 µL	800 µL	1000 µL
500 µg	35 %	60%	70%	75%	75%	75%
	2.5 µg/µL	2.3 µg/µL	1.2 µg/µL	0.8 µg/µL	0.6 µg/µL	0.5 µg/µL
250 µg	40%	65%	75%	80%	80%	80%
	1.9 µg/µL	1.1 µg/µL	0.6 µg/µL	0.4 µg/µL	0.3 µg/µL	0.2 µg/µL
100 µg	45%	70%	80%	85%	85%	85%
	0.7 µg/µL	0.4 µg/µL	0.2 µg/µL	0.1 µg/µL	0.1 µg/µL	0.1 µg/µL
50 µg	30%	75%	85%	90%	90%	90%
	0.3 µg/µL	0.2 µg/µL	0.1 µg/µL	0.1 µg/µL	0.1 µg/µL	< 0.1 µg/ µL

DNA recovery
DNA concentration

Table 2.1: Table from Nucleobond® Xtra Midi EF Kit, showing the DNA recovery for each elution volume, determined by the amount of DNA measured with the Nanodrop, before the isopropanol step.

Table 2.1 - DNA recovery for different volumes of elution given by the Nucleobond Xtra Midi EF Kit

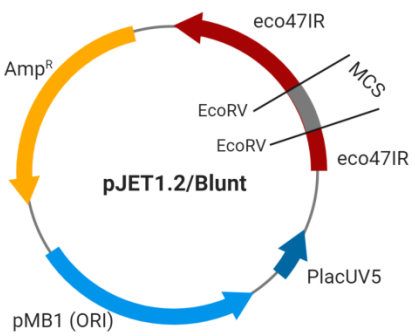
2.4.6 - BACTERIAL STOCK

After the overnight 50ml bacterial culture for midiprep (section 2.4.5), 500µl of bacterial culture were added to 500µl of 50% (v/v) glycerol and directly frozen at -80°C.

2.5 MOLECULAR BIOLOGY

2.5.1 - PLASMIDS

All the plasmids used during the project are listed in table 2.2. The first column corresponds to the name of the plasmids with their origin. The second column corresponds to some informations about the plasmid function and the third column corresponds to a graphic representation of each plasmid.

pJET1.2/Blunt (ThermoFisher Scientific)	Plasmid for cloning. Contains the lethal gene eco47IR disrupted by the EcoRV blunt cut. This gene allows positive selection for the cloning. If there is no insert, the gene will ligate and create a restored eco47IR gene, which will kill the bacteria.	

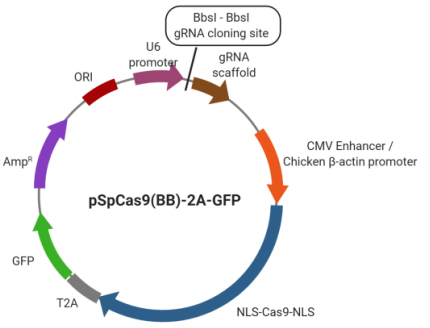
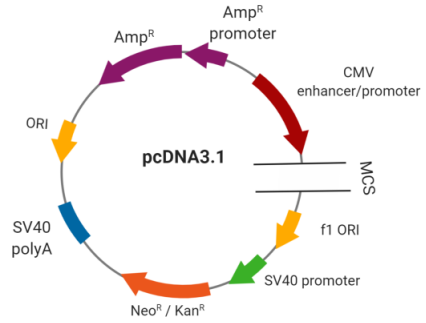
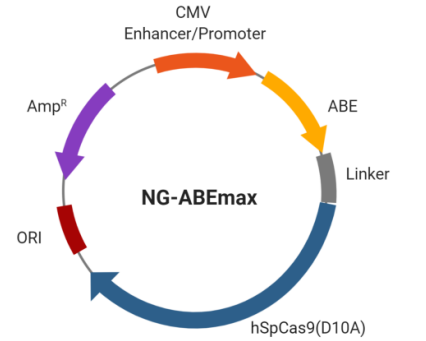
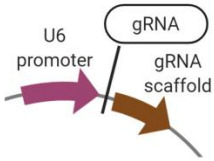
pSpCas9(BB)-2A-GFP (Addgene)	Plasmid for sgRNA cloning, containing the gRNA backbone. Plasmid also expressing SpCas9. Allows co-expression of the sgRNA and Cas9 with single plasmid.	
pcDNA3.1(+)	Empty expression plasmid used as a negative control for experiments.	
Plasmid NG-ABEmax (Addgene)	Plasmid expressing the nSpCas9-NG(D10A), recognising the PAM 5'-NG-3', with the ABEmax deaminase fused to it.	
Plasmid gRNA ABE (IDT)	gRNA plasmids for base editing were designed as shown in the picture and were ordered from IDT to be cloned in the pUCIDT(Amp) cloning plasmid.	 cloned into the pUCIDT(Amp) plasmid from IDT

Table 2.2: Plasmid constructs and respective details

Table 2.2 - Plasmid constructs and respective details

2.5.2 - MOLECULAR BIOLOGY REAGENTS

Electrophoresis

- Tris-Acetate-EDTA (TAE) Buffer 50X stock solution: 242g of Tris Base (Sigma-Aldrich), 57.1ml of glacial acetic acid (Sigma) and 100ml of 0.5M EDTA (Sigma) were added to up to 1L of deionised water.
- Agarose (Sigma-Aldrich)
- SafeView Nucleic Acid Stain (NBS Biologicals)
- UVIpro™ Gel Documentation System (UVITEC)
- 100bp DNA Ladder (NEB)

DNA assembly

- All restrictions enzymes and their Buffers were ordered from New England Biolabs (NEB)
- T4 ligase and T4 ligase buffer (NEB)
- Thermosensitive Alkaline Phosphatase (Promega)
- NEB Buffer 2 (NEB)

Polymerase Chain Reaction (PCR) reagents:

- GoTaq® Flexi DNA Polymerase containing 5X Green GoTaq® Flexi Buffer, 25mM $MgCl_2$ and the GoTaq® Flexi DNA Polymerase (Promega).
- Deoxynucleotide (dNTP) Solution Mix (NEB)
- All primers were purchased on Eurofins and diluted at a working concentration of 10µM.

CRISPR guides oligonucleotides:

- Every oligonucleotide for gRNA cloning were ordered on Eurofins.

CRISPR proteins and RNA guides for Ribonucleoproteins:

- Alt-R® S.p. HiFi Cas9 Nuclease V3 (Integrated DNA Technologies; IDT)
- Alt-R® CRISPR-Cas9 crRNA (Integrated DNA Technologies; IDT)
- CRISPRevolution sgRNA EZ Kit (Synthego). Contains a tube of sgRNA with 2'-O-methyl 3' phosphorothioate modifications in the first and last 3nucleotides, a tube of Tris-EDTA Buffer and a tube of Nuclease-free water.
- Alt-R® CRISPR-Cas9 tracrRNA (IDT)
- Alt-R® A.s. Cas12a (Cpf1) V3 (IDT)

- Alt-R® CRISPR-Cpf1 crRNA (IDT)

Other Reagents:

- IDTE Nuclease free buffer (IDT)
- The nuclease-free water was taken from the RNA extraction kit (section 2.5.7a) and the endotoxin-free plasmid DNA purification kit (section 2.4.5b-2).

2.5.3 - gRNA DESIGN FOR spCAS9

To find a guide sequence in a genome of interest for SpCas9, the software CHOPCHOP v3 was used (178). In this software, the DNA sequence of interest was pasted in the "target" box, the species used was "*homo sapiens"*, the protein used was CRISPR/Cas9 and the option used was Knock-In. The software analysed the sequence for guide sites and sequence specificity, then, it gives a rank of guides from the most specific to the less specific and most efficient to less efficient. To choose the guides of interest, 3 guides ranked in the top 20 were ordered and tested to check for their efficiency (see section 2.5.8). The most efficient guide was used for further experiments.

2.5.3a - gRNA oligonucleotides design for cloning in the pSpCas9(BB)-2A-GFP plasmid

Once the guide sequence was chosen, the 20 nucleotides before the PAM were ordered in the orientation 5'→3', containing the sequence "CACC" followed by a guanine upstream the spacer sequence at the 5' end. The "CACC" sequence was used for cloning with the BbsI enzyme in the pSPCas9(BB)-2A-GFP plasmid, and the guanine was added for the first base of the U6 transcript (179).

This designed oligonucleotide was called the top strand guide. To design the bottom strand, the reverse complementary sequence of the 20 nucleotides was designed (to order a 5'→3' orientation DNA) and the sequence "AAAC" was added at the 5' end for BbsI cloning. A cytosine was added at the 3' end for the U6 transcript.

2.5.3b - gRNA molecules design

2.5.3.b-i crRNA molecules for Cas9 protein

After finding the guide sequence using the software CHOPCHOP v3 (178) , the 20 nucleotides upstream the NGG PAM were entered on IDT or Synthego website in the 5' to 3' orientation.

2.5.3.b-ii gRNA molecules for Cas12a protein

For Cas12a guides, the sgRNA used in chapter 3 were designed manually. The two closest TTTV PAM sites around the W1282X *CFTR* mutation were used for design and testing. The 20 nucleotides downstream the TTTV PAM were entered in IDT website in the 5' to 3' orientation.

2.5.3.b-iii gRNA plasmid design for base-editing

The guide RNA for base editing were ordered as minigene plasmids from IDT. The guides were manually designed to have the U6 promoter, the 20nt gRNA spacer sequence and the gRNA SpCas9 backbone.

2.5.3c - Annealing of the crRNA and the trans-activating crRNA (tracrRNA)

To anneal the crRNA and tracrRNA, 4µl of 100µM designed crRNA oligonucleotides and 4µl of 100µM tracrRNA oligonucleotides were mixed with 2µl of IDTE Nuclease free buffer to a final concentration of 40µM. The mixture was heated at 95°C for 5min and then left on the bench to cool down until it reached room temperature.

2.5.4 - DNA MOLECULES ASSEMBLAGE

2.5.4a - Enzymatic digestion

In the case where two pieces of DNA were into two different vectors and needed to be assembled together, the first step was to cut those both vectors with two compatible enzymes, located around the sequences of interest. Having two compatible enzymes, means having two enzymes that have their maximum efficiencies in the same buffer.

The cutting map of the vectors were analysed using the Ape software. Enzymes cutting around the sequences of interest were picked and checked on NEB that two enzymes had their highest efficiencies in the same buffer.

In an Eppendorf tube, one microgram of DNA was digested with 1µl of each enzyme, in 2µl of the appropriate buffer and up to 20µl of nuclease free water. The reaction was set at 37°C for 2 hours.

2.5.4b - Electrophoresis

To prepare a 1% (w/v) agarose gel, 60mg (for a small gel (10 wells)) or 120mg (for a big gel (12 to 20 wells)) of agarose was mixed with 60ml or 120ml 1X TAE Buffer respectively. The mix was boiled in the microwave until the agarose was completely dissolved. Once the mix had cooled down to about 60°C, 50nl/ml of SafeView were added. The mix was poured into a casting tray and left to cool down at room temperature until solidified. The gel was then added to an electrophoresis tank filled with 1X TAE Buffer and containing 50nl/ml of SafeView. The samples and ladder (section 2.5.2) were loaded into the wells. The gel was run at 120V for 30min to 1 hour depending on how well the bands needed to be separated. The bands were observed by transillumination. If the bands of interest needed to be extracted from the gel, the UV was turned on and, using a blade, the bands were cut out of the gel and added to an Eppendorf tube. A UV protector mask was used to protect the face and eyes from UV. The ladders used were the 100bp

2.5.4c - DNA gel extraction

The DNA extraction from agarose gels was performed using the High Pure PCR Product Purification Kit (Roche), and following the kit's protocol.

The extracted bands were weighed and for every 100mg of gel, 300µl of binding buffer were added to the tube. The mix was incubated at 56°C for 10min, until the gel was completely dissolved. One hundred fifty microlitres of isopropanol for every 150mg of gel were added to the tube, and the content was poured into a High Pure filter tube. The column was centrifuged for 1min at maximum speed, and 500µl of wash buffer were added to the filter. After another run of centrifugation, 200µl of wash buffer were added and the tubes were centrifuged for 1min at maximum speed. The elution of the DNA was done using 50µl of elution buffer and centrifuged for 1min at maximum speed.

2.5.4d - Alkaline phosphatase treatment

From the 50µl of gel extracted DNA (section 2.5.4c), 35µl were treated with 1µl of TSAP enzyme and 1X of the MultiCore 10X buffer (4µl). The reaction was incubated at 37°C for 20min and the enzyme was heat-inactivated at 74°C for 15min.

2.5.4e - Ligation

To ligate pieces of DNA together, a ratio 3:1 (inserts:vectors) were mixed together with T4 ligase and its T4 10X Buffer to a 1X final concentration. The ratios were calculated using the NEBioCalculator tool (180). As a negative control, the vector alone was mixed with the ligase and buffer. The mixes were incubated for 15min at RT and the ligase was inactivated 10min at 65°C.

2.5.4f - pJET blunt end cloning

The assemblage of DNA into a pJET2.1 plasmid was done using the CloneJET PCR Cloning Kit (Thermo Scientific) and following the kit's protocol. The CloneJET kit allows assembly of DNA with blunt ends and phosphate groups in their 5' ends.

In the case where the DNA contains blunt ends (e.g., gBlocks® Gene Fragments, IDT), 0.15pmol ends of blunt ends DNA fragments were mixed with a 1X final concentration of 2X Reaction Buffer, 0.05pmol ends of pJet1.2/Blunt Cloning Vector and 1µl of T4 DNA ligase in up to 20µl of nuclease-free water. The reaction was incubated at RT for 15min.
In the case where the DNA contains overhangs, 0.15pmol ends of sticky ends DNA fragments were mixed with 10µl of 2X Reaction Buffer, 1µl of DNA blunting enzyme and up to 18µl of nuclease-free water. The reaction was incubated at 70°C for 5min and chilled on ice, before adding 0.05pmol ends of pJet1.2/Blunt Cloning Vector to the reaction along with 1µl of T4 DNA ligase. The reaction was incubated at RT for 15min.

2.5.4g - gRNA oligonucleotides annealing

In an Eppendorf tube, 100pmol of the top strand oligonucleotide and 100pmol of the bottom strand oligonucleotide (section 2.5.3a) were added to 16µl of nuclease-free water and 2µl of 10X NEB buffer 2. The oligonucleotides were denatured by placing the tube in a heat block for 5min at 95°C, and the block was placed on a bench until the mixture reached room temperature, to allow slow hybridisation of the oligonucleotides.

2.5.4h - Golden gate cloning for sgRNA production using pSpCas9(BB)-2A-GFP vector

In a 0.5ml PCR tube, 100ng of the pSpCas9(BB)-2A-GFP plasmid was mixed with 1µl of BbsI restriction enzyme, 1µl of 100µM annealed gRNA oligonucleotides (see section 2.5.4g), 1µl of T4 ligase, and 2µl of 10x T4 ligase buffer in up to 20µl total reaction volume with nuclease-free water.

For subsequent cloning of those assembled DNA, 5µl of the reaction mixes were used for transformation and plasmid amplification (sections 2.4.3, 2.4.4 and 2.4.5).

2.5.5 - DNA EXTRACTION

2.5.5a - DNA extraction for a large number of cells

The DNA extraction was performed using the DNeasy Blood and Tissue Kit (Qiagen), following the kit's protocol. This kit was used for cells from a 24-well, 12-well or 6-well plates. Cells were collected by removing the medium from the wells, washing with PBS (HEK cells) or trypsin (HBE cells), and adding trypsin to cover all the well. The plate was incubated at 37°C and once the cells were detached, the trypsin was inactivated with medium and put in an Eppendorf tube. The tube was centrifuged for 5min at 300g and the pellet was resuspended with 200µl of 1X PBS and 20µl of proteinase K. Two hundred microlitres of buffer AL were added for cell lysis and the tube was incubated at 56°C for 10min. Two hundred microlitres of ethanol 96% (v/v) were added to the lysate, and the mixture was added to a DNeasy Mini spin column. The column was centrifuged for 1min at 6,000g. Two steps of wash were done using 500µl of buffers AW1 and AW2 subsequently. The buffer AW2 was centrifuged for 3min to dry the column of the remaining ethanol. The elution step was done by centrifuging 200µl of buffer AE at 6,000g for 1min.

2.5.5b - DNA extraction for a small number of cells

The DNA extraction has been performed using the QuickExtract™ DNA extraction solution (Lucigen). This kit was used for cells from 96-well plates. The cells were trypsinised like in section 2.5.5a and resuspended in 200µl of medium. The suspensions were added in a 96-wells V-bottom PCR plate and centrifuged for 5min at 300g. The pellets were resuspended with 20µl of QuickExtract solution. The cell suspensions were transferred to PCR tubes and incubated in a PCR machine 15min at 65°C and 10min at 98°C.

2.5.6 - POLYMERASE CHAIN REACTION (PCR)

2.5.6a - PCR primers

Primers	
Name	sequence 5' → 3'
3849+10kb FW	TCATTCAGTGGGTATAAGCAGCA
Superexon 5'REV	TATCACTCCAAAGGCTTTCCTCCAC
CFTR exon 21 FW	ACTCCAGCATAGATGTGGATAGC
end intron 22 REV	ACTTCAATGCACCTCCTCCCTG
modified in22 FW	GCAAGTCTGCTACAACGTATGACTG
Int22 -800bp FW	CTCAAAGCACCCTCCTCAAGTTC
Int22 -700bp FW	CTAGCACTTGATCTCCTCATGCAG
in22 FW EMGin21-in24	TCTGTTCCAAGGTTGTTTGTCTCC
exon 23 codon opt REV	CTCACGCCGTCAATCTGAATCTC
in23 REV	TCGCAAAGCATTTCTCAACCTGG
exon 23 FW	TGGGAAGAACTGGATCAGGGAAG
int23 REV EMGin21-in24	CAGGCCACTGAAGCAGGAAC
intron 23 REV	TCGCAAAGCATTTCTCAACCTGG
junction 24-25 CFTR REV	CCCAACCTCATCTGCAACTTTCC
5' integration SEv2 FW	TTGATTTCTGGAGACCACAAGGT
5' integration SEv2 REV	TATCACTCCAAAGGCTTTCCTCCAC
NGS 5' FW	TCATCTTGATTTCTGGAGACCACA
NGS 5' REV	GGTGCTAGCTGTAATTGCATTGT
backbone FW SEv2	GCGATTAAGTTGGGTAACGCCAG
GAPDH FW	ACCCACTCCTCCACCTTTGA
GAPDH REV	CTGTTGCTGTAGCCAAATTCGT

Table 2.3: Primers' sequences

Table 2.3 - Primers' sequences

2.5.6b - PCR protocol

In PCR tubes, between 500 and 1,000ng of genomic DNA (100ng for purified DNA) were used. In a 0.5ml tube, a mastermix was made containing all the reagents in the table below, except DNA.

Reagent	Volume (µl) per tube	
5X Green GoTaq® Flexi Buffer	5µl	
$MgCl_2$	2.5µl	
dNTP	1µl	
Primer forward	0.5µl	
Primer reverse	0.5µl	
GoTaq® Flexi DNA Polymerase	0.25µl	
Nuclease-free H_20	5.25µl	10.25µl
DNA (500-1000ng)	10µl	5µl
Total Volume	**25µl**	

In each PCR tube, up to 25µl of mastermix were added to the DNA (e.g. 15µl of mastermix were added if 10µl of DNA were used). The tubes were vortexed and spun down before incubated in the PCR machine.

PCR program:

The PCR program used is shown in Figure 2.1.

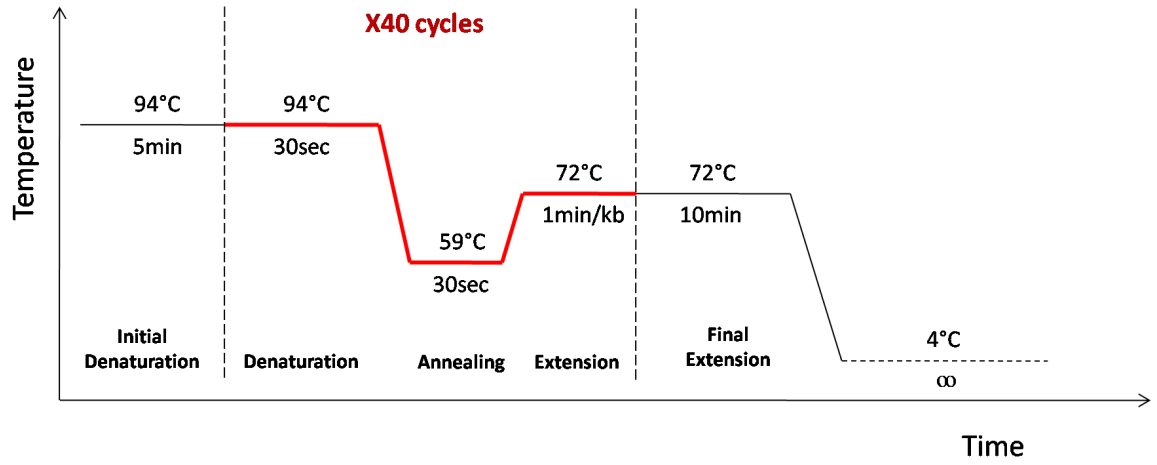


Figure 2.1: PCR program. All the PCR were performed using the same program and temperatures. The extension time was different in function of the size of the amplicon.

Figure 2.1 - PCR program

After PCR, the DNA was analysed by gel electrophoresis gel to confirm the success of the amplification (section 2.5.4b).

2.5.6c - Extraction of the PCR product from the PCR mix

The extraction was done using the MinElute® PCR Purification Kit (Qiagen) and following the kit's protocol. Five volumes of Buffer PB were added to 1 volume of PCR reaction. The mix was added to a MinElute column and centrifuged at maximum speed for 1min. The column was washed with 750µl of Buffer PE and centrifuged using the same conditions. The flow through was discarded and the emptied column was centrifuged 1min at maximum speed to remove any remaining ethanol. The DNA was finally eluted using 20µl of Buffer EB and centrifuged for 1min at maximum speed.

2.5.7 - REVERSE TRANSCRIPTION OF CELLULAR RNA

2.5.7a - RNA extraction

RNA extraction was done using the NucleoSpin® RNA Kit (Macherey-Nagel) and a modified version of its protocol. Cells were harvested and pelleted like in section 2.5.5a. To lysate the cells, 350µl of buffer RA1 and 3.5µl of β-mercaptoethanol were added to the cell pellet, following by 350µl of 70% (v/v) ethanol. The samples were added to a NucleoSpin® RNA column and centrifuged for 30sec at 11000g. Three hundred fifty microlitres of MDB buffer were added to the column and the samples were centrifuged for 1min at 11000g.
Three washing steps were done using 200µl of buffer RAW2, 600µL of buffer RA3 and 250µl of buffer RA3 respectively. The RNA was eluted using 40µl of RNase-free water.

RNA concentrations were measured at this step, using a Nanodrop.

2.5.7b - DNase treatment

The DNase treatment was done using TURBO DNA-*free*™ Kit (Invitrogen). Five microlitres of 10X TURBO DNase Buffer were added to 35µl of RNA sample, along with 1µl of TURBO DNase. The reactions were incubated for 45min at 37°C. Another microlitre was added to the reaction before incubating again for 45min at 37°C.

For enzyme deactivation, 10µl of inactivating reagent were added to the reaction and incubated 2min at RT vortexing frequently. The reaction was centrifuged for 1.5min at 16249g and the RNA containing supernatants were collected.

2.5.7c - Reverse transcription (RT)

The **R**everse **T**ranscriptions (RTs) were done using the RevertAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher). On ice, 1µg of RNA in up to 5.5µl of nuclease-free water was prepared in a PCR tube for each condition. In a mastermix tube, 0.5µl per tube of oligodT and 1µl per tube of dNTP were added. One point five microlitres of the mastermix were added to the RNA. The reaction was incubated for 5min at 65°C and chilled on ice. In another mastermix tube, 2µl per tube of Reaction Buffer, 0.5µl per tube of Ribolock enzyme and 0.5µl per tube of MLVRT enzyme were prepared. Three microlitres of this mix were added to the RNA.

The tubes were incubated in a thermocycler for 1h at 42°C and for 5min at 70°C. After the RT, 20µl of nuclease-free water were added to the reaction to dilute the RNA 1:3.

2.5.8 - SANGER SEQUENCING ANALYSES

2.5.8a - Knock-in efficiency and indels formation

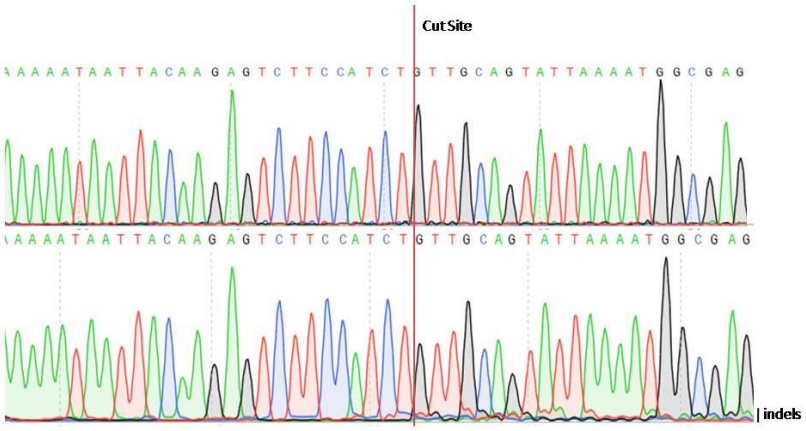
To analyse the indels efficiencies of a CRISPR protein and its guides, or to analyse the efficiency of integration of an external DNA molecule in the genome, Sanger sequencings were performed. First, a transfection was completed using an empty plasmid as control (pcDNA3.1; section 2.5.1) and the different conditions to analyse (transfection methods in section 2.3). After DNA extraction (see section 2.5.5a), a PCR was performed around the cut site (see section 2.5.6) and the PCR product was sent to Eurofins for Sanger sequencing (Figure 2.2A). The Sanger sequencing results in a chromatogram where each peak corresponds to a nucleotide present in the sequence. In the case where there are several different sequences in the mix, there will be superposed peaks where the size depends on the frequency of each nucleotide.

The Deskgen analysis tool "**T**racking of **I**ndels by **De**composition" (TIDE) (181) uses the Sanger sequencing data files to give an approximation of the number of indels by comparing a control chromatogram with the experimental chromatogram (Figure 2.2B). The Sanger sequencing chromatogram (.ab1 file) from the pcDNA (control) transfected cells and the experimental conditions are added to the programme, which will use an algorithm to estimate

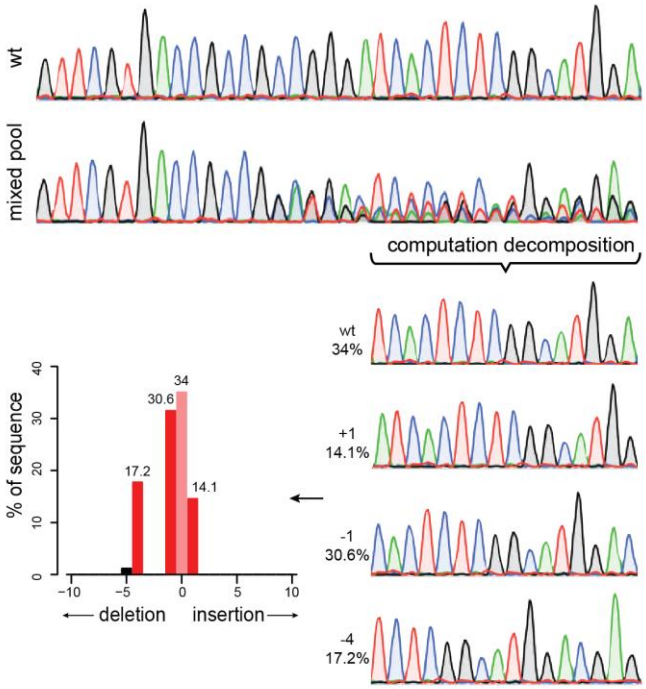
the number and types of indels created by CRISPR/Cas9. For each type of indels, a p-value will be associated for the significance of each of those indel.

The Synthego analysis tool "**I**nference of **C**RISPR **E**dits" (ICE) (182) uses the Sanger sequencing data files to give an approximation of the number and types of indels and the efficiency of knock-in, by comparing the pcDNA control chromatogram with the experimental chromatograms. Moreover, this analysis tool can also give an estimation of the sequence for each indel present in the sequence (Figure 2.2C).

A)



B)



C)

Figure 2.2: Indels and knock-in efficiency analyses for genome editing. A) Example of a Sanger sequencing data file showing a control sequence (top) and a sequence containing indels (bottom). B) Principle for Deskgen "TIDE" analysis. An algorithm for decomposition is used to estimate the numbers of indels in a mixed population ***(figure from tide.nki.nl*** *(181))**. C) Example of a Synthego "ICE" analysis from a Sanger sequencing profile. The upper panel shows a comparison of indels efficiencies for two different Cas12a gRNAs. The profile showed that crRNA1 was the most efficient one. The bottom panel shows an example of a list of the different types of mutations the ICE algorithm gave for a Cas12a HDR experiment. The list showed the contribution of each indel in the DNA mix.*

Figure 2.2 - Indels and knock-in efficiency analyses for genome editing using A) Sanger sequencing, B) TIDE and C) ICE analysis tools

2.5.8b - Base editing efficiency analyses

Base editing efficiencies where determined using the EditR software (183). EditR is a software using an algorithm to estimate the percentage of base editing efficiency from a Sanger sequencing. The results are shown as a table giving the percentage of each base in the chromatogram.

2.6 NEXT GENERATION SEQUENCING FOR PCR AMPLICONS

2.6.1 - SAMPLES PREPARATION AND ANALYSIS

Five hundred nanograms of PCR purified amplicons were sent to the GENEWIZ company for Amplicon-EZ deep sequencing. Amplicons of 150 to 500bp are analysed through Illumina sequencer with 50,000 reads per sample. The adaptor sequences used by Genewiz are:

Forward read	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
Reverse read	5′-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′

Raw data are sent in compressed fastq files.

2.6.2 - DATA ANALYSIS

Data were decompressed and analysed using the QIAGEN CLC Genomics Workbench software.

2.7 FUNCTIONAL ASSAYS

2.7.1 - FACS ANALYSIS AND CELL SORTING

All **F**luorescence-**A**ctivated **C**ell **S**orting (FACS) analyses and cell sorting were done using the BD FACS Aria™ Fusion cell sorter and were kindly performed by Panagiota Stamou, from the **A**limentary **P**harmabiotic **C**entre (APC) Microbiome Ireland Research Centre in University College Cork. Cells were collected by removing the medium from the wells, washing with PBS (HEK cells) or trypsin (HBE cells), and adding trypsin to cover all the cells in the well. The plate was incubated at 37°C and once the cells were detached, the trypsin was inactivated with medium. The detached cells were centrifuged at 500rpm for 5min. The cell pellet was re-suspended with PBS 1X + 2% (v/v) FBS, through a 70µm filter that prevent cell clamps. The tubes were stored on ice until analysis.

The FACS analyses were done as shown in Figure 2.3.

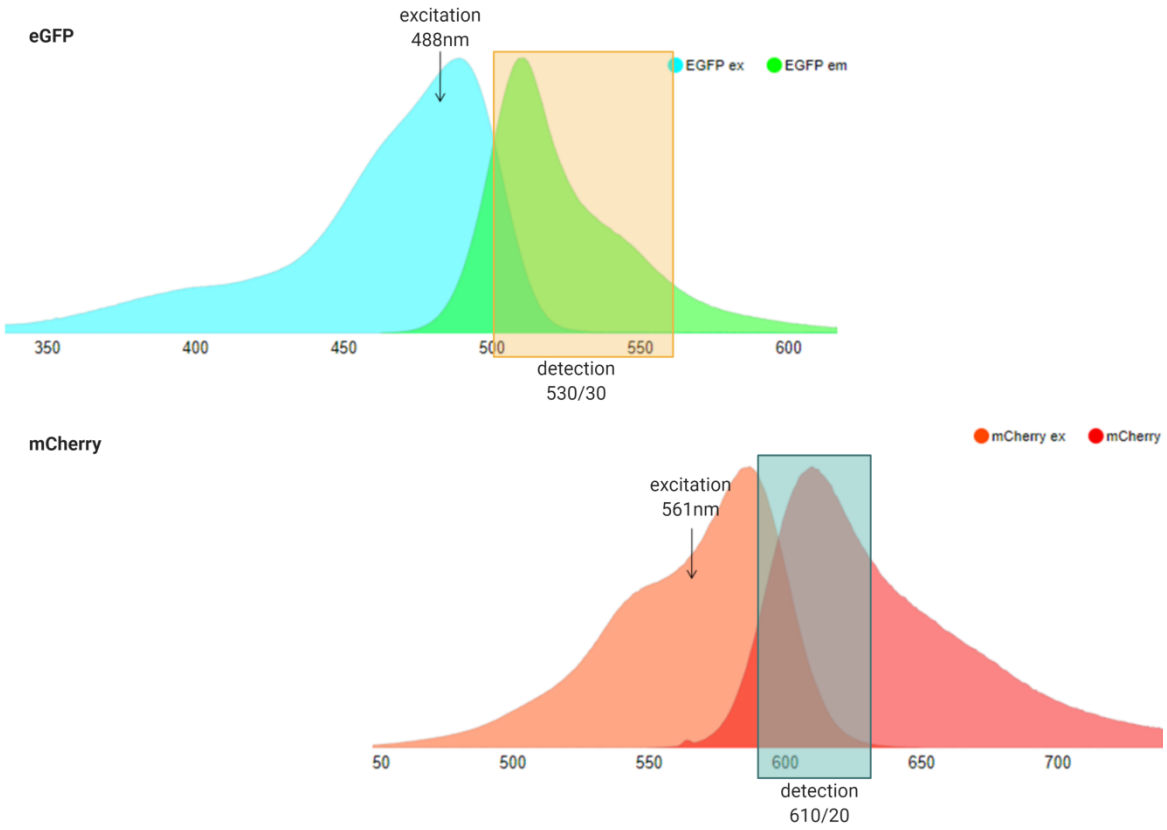


Figure 2.3: Excitation and emission spectra for both GFP and mCherry proteins. The GFP protein was excited using a 488nm laser and was detected with a 530/30 band pass filter. The mCherry protein was excited using a 561nm laser and was detected with a 610/20 band pass filter.

Figure 2.3 - Excitation and emission spectra for both GFP and mCherry proteins

Using lasers and filters available in the cell sorter, GFP was excited using the 488nm laser and mCherry using the 561nm laser. The detections were done using the 530/30 band pass and 610/20 band pass filters respectively.

2.7.2 - RESISTANT MONOLAYER FORMATION FOR USSING CHAMBER

In snapwells^{TM}, 2x10^{5} cells per well were plated to make a liquid-liquid culture interface. After 7 days in culture, the **T**rans-**E**pithelial **E**lectrical **R**esistance (TEER) was measured using an **E**pithelial **V**olt**O**hm**M**eter (EVOM).

Measuring the TEER enables identification of the monolayer resistance. Tight junctions between cells will give a high electrical resistance, however, if the epithelium is leaky, the TEER will be low and further functional assays using Ussing chamber will not be possible.

CHAPTER 3 - PRECISE GENOME EDITING: CORRECTING A SINGLE SPECIFIC MUTATION BY HDR

Gene editing for Cystic Fibrosis has been attractive for many years with different groups trying to develop techniques correcting the most common CF mutation, F508del (184). The purpose of this PhD project was to develop and compare different efficient gene editing techniques used in other diseases, in order to optimise genetic correction in CF for potential future therapeutic prospects.

Because F508del can be treated with drugs, the focus here is on four different gene editing techniques used in parallel to correct a common class I CF mutation, W1282X. This mutation cannot be treated with the currently available drugs. Those four different techniques use either the Homology-Directed Repair pathway (techniques described in chapter 3 and chapter 6), the Non-Homologous End Joining repair pathway (described in chapter 5) or the Mismatch Repair pathway (described in chapter 4), and can correct either a unique mutation (techniques described in chapters 3 and 4) or many mutations at the time (techniques described in chapters 5 and 6). The purpose of this project will be to compare the techniques and display their advantages and limitations, in order to allow researchers to have a clear overview of gene editing for CF research.

W1282X is the 6th most common CF mutation, concerning 1.2% of CF-patients, and is localised in the exon 23 of the *CFTR* gene. The mutation is a nonsense mutation, originating from a substitution from an adenine (A) to a guanine (G), creating a premature stop codon 5'-TGA-3'. This stop codon creates mRNA instability, where the premature stop codon in the mRNA is recognised by the nonsense mediated decay (NMD) pathway, leading to degradation of the mRNA. Consequently, there are few to no functional CFTR protein molecules at the cell surface.

The central hypothesis of this project is that each technique will be able to correct the W1282X mutation in the *CFTR* gene, restore mRNA stability and CFTR protein function, all of this with an efficiency relevant for CF.

In this chapter, the correction of the W1282X mutation in exon 23 of the *CFTR* gene was attempted using the CRISPR/Cas12a system and the HDR pathway. The purpose was to compare efficiencies with the CRISPR/Cas9 editing system. Because the experiments using Cas9 was done by another PhD student in the laboratory, the Cas9 data will not be shown, but will be discussed.

3.1 GUIDE DESIGN AND VERIFICATION

Because HDR efficiency is higher when the modifications to make are within 20bp from the cut site (101), the first step was to design a guide cutting at a good efficiency around the 20bp flanking the W1282X mutation. The Cas12a PAM site being a 5'-TTTV-3' sequence, all the PAM sequences localised around the W1282X mutation, in either the coding or non-coding strand, were identified. The two gRNAs making a DSB the closest to the W1282X mutation were designed and tested for efficiency. The guide 1 (G#1) is localised on the transcribed strand and makes a DSB 9bp downstream of the mutation. The guide 2 (G#2) is localised on the non-transcribed strand and cut 20bp upstream of the mutation (Figure 3.1).

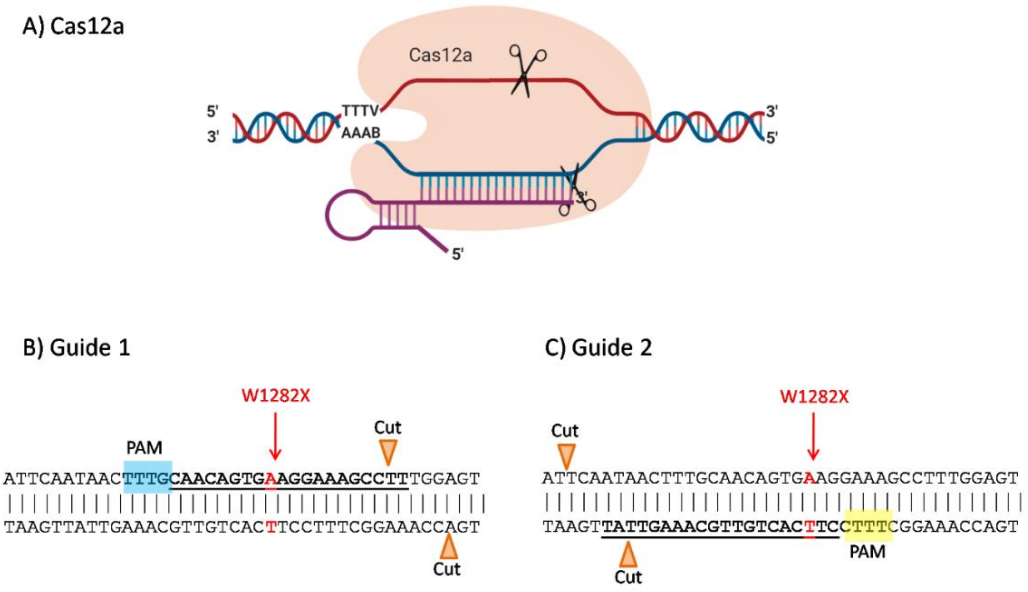


Figure 3.1: gRNA design for W1282X correction using Cas12a and HDR. A) AsCas12a mechanism (described in section 1.8.2). B) and C) show the design of Guide 1 and Guide 2 respectively, relative to the W1282X mutation. The coloured squares show the PAM sequence, the underlined sequence corresponds to the spacer and the orange arrows show the cut sites. Finally, in red is shown the position of the W1282X mutation.

Figure 3.1 - gRNA design for W1282X correction using Cas12a and HDR

To verify the guides' efficiency, an electroporation of the ribonucleoprotein (RNP), containing one of each guide with the Cas12a protein, was done in 16HBE14o$^-$ containing the W1282X mutation. The efficiency was measured by the indels frequency achieved by each RNP. Four days post-transfection, for each condition of electroporation, the DNA was extracted, and a PCR was performed using a set of primers targeting exon 23 (exon23 FW) and intron 23 (intron23 REV). The amplicons were purified and sent for Sanger sequencing. The TIDE (**T**racking of **I**ndels by **DE**composition) analysing tool from DESKGEN was used to determine, which guide produced the more indels (Figure 3.2).

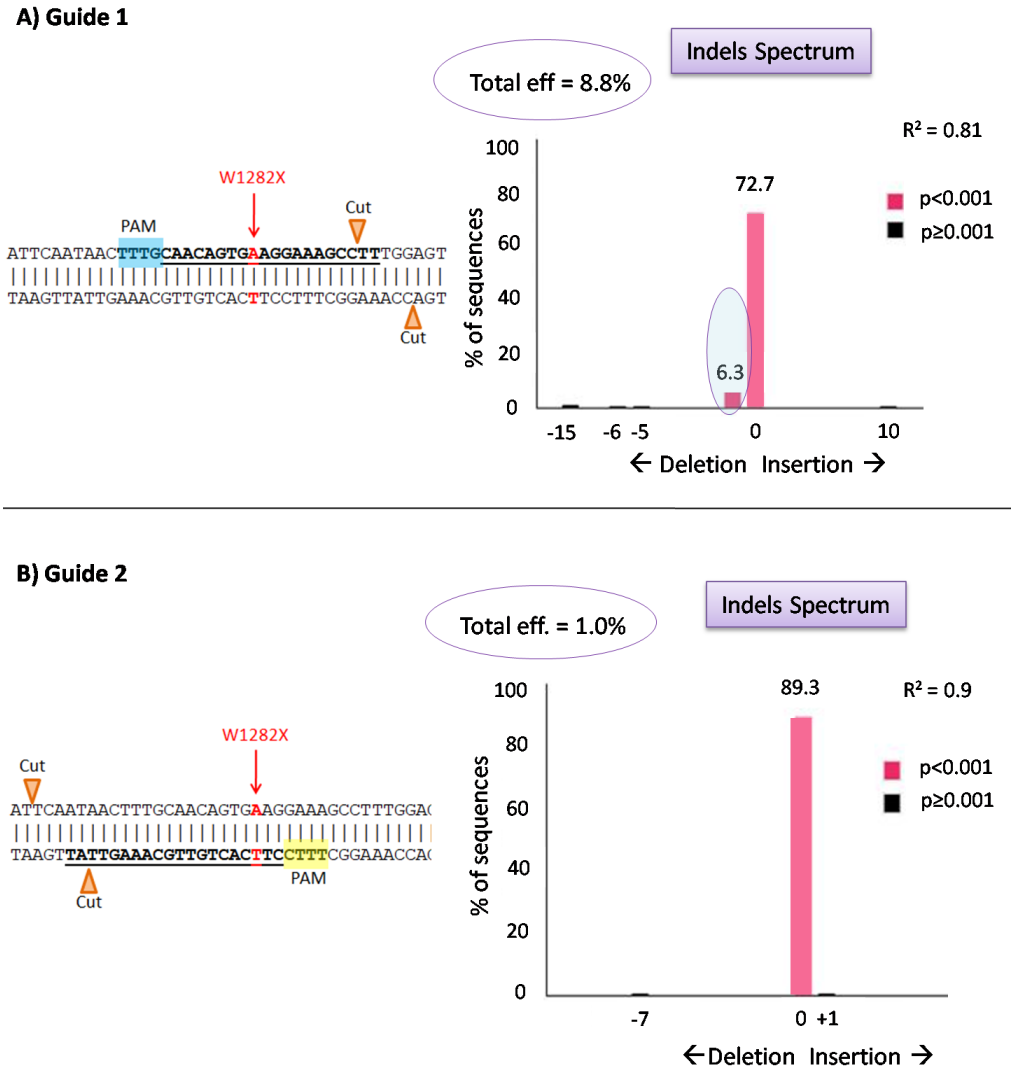


Figure 3.2: Indels frequency for each W1282X gRNA. An electroporation was performed with the AsCas12a and Guide 1 or Guide 2. Four days post-transfection the DNA was extracted, and a PCR was performed around the cut site. The amplicons were sent for sequencing and the TIDE analysing tool was used to determine the percentage of editing efficiency for each gRNA. A) Guide 1: The total efficiency was around 9%. B) Guide 2: The total efficiency was around 1%.
Bars in black represent percentages with a low p-value (≥0.001) considered not significant.

Figure 3.2 - Indels frequencies for each W1282X gRNA using AsCas12a.

The results of the TIDE analysis showed a total editing efficiency of about 9% for Guide 1 and 1% for Guide 2, reflected by the indels frequency. Those results demonstrate that Guide 1 is the most efficient gRNA to cut close to the W1282X mutation. Importantly, the 8.8% efficiency observed, was analysed from a pool of transfected and non-transfected cells. Moreover, as described in the introduction (section 1.9.1), the NHEJ pathway often results

in perfect resection, mostly when the DSB generates blunt ends or compatible ends, like Cas9 or Cas12a respectively.

Therefore, taking into account the efficiency of transfection as well as the precise resection from the cells, it can be expected that the efficiency of cutting is higher than those ~9%. However, because neither the guide nor the protein has a fluorescent or an antibiotic marker, the transfection efficiency could not be assessed.

Considering all those data, the Guide 1 was the gRNA used for further experiments.

3.2 HDR DONOR TEMPLATE DESIGN

In order to perform HDR and correct the W1282X mutation, the design of an optimal donor template is important to achieve the best efficiency. In 2018, a comparison of editing efficiencies between different Cas9 and Cas12a proteins, using different single strand oligodeoxynucleotides (ssODN) donors, was published (115).. The conclusion was that AsCpf1 (AsCas12a) and Cas9 were editing genes at a similar efficiency and AsCpf1 performed more efficiently using an asymmetric ssODN donor template with a longer PAM-proximal homology arm of about 77nt and a shorter PAM-distal homology arm of 37nt (77/37 donor). Moreover, the efficiency was higher when the sequence of the donor was the same as the non-target strand sequence (Figure 3.3; (115)).

For SpCas9 nuclease, the optimal HDR donor was an asymmetric donor with a shorter PAM-proximal homology arm of 37nt and a longer PAM-distal homology arm of 77nt (37/77 donor). Moreover, the efficiency was higher when the sequence of the donor was the same as the target strand sequence (115).

The design of the ssODN donor, for W1282X correction, was performed based on data from that study. Two ssODN donor templates were designed. Both donor sequences had the non-target strand sequence and contained both the correction of the mutation W1282X (A → G) and a silent mutation on the PAM site (T→C) to disrupt the gRNA recognition site after recombination (Figure 3.3). Disrupting the PAM site is important to prevent Cas12a from cutting again after recombination and creating unwanted indels. One of the donors contained a 77nt long PAM-proximal homology arm and a 37nt shorter PAM-distal homology arm. This donor was called 77/37 donor, and it was the donor with the highest efficiency for Cas12a, described previously.

The second donor had a 37nt shorter PAM-proximal homology arm and a 77nt longer PAM-distal homology arm. This ssODN was called 37/77 donor (Figure 3.3).

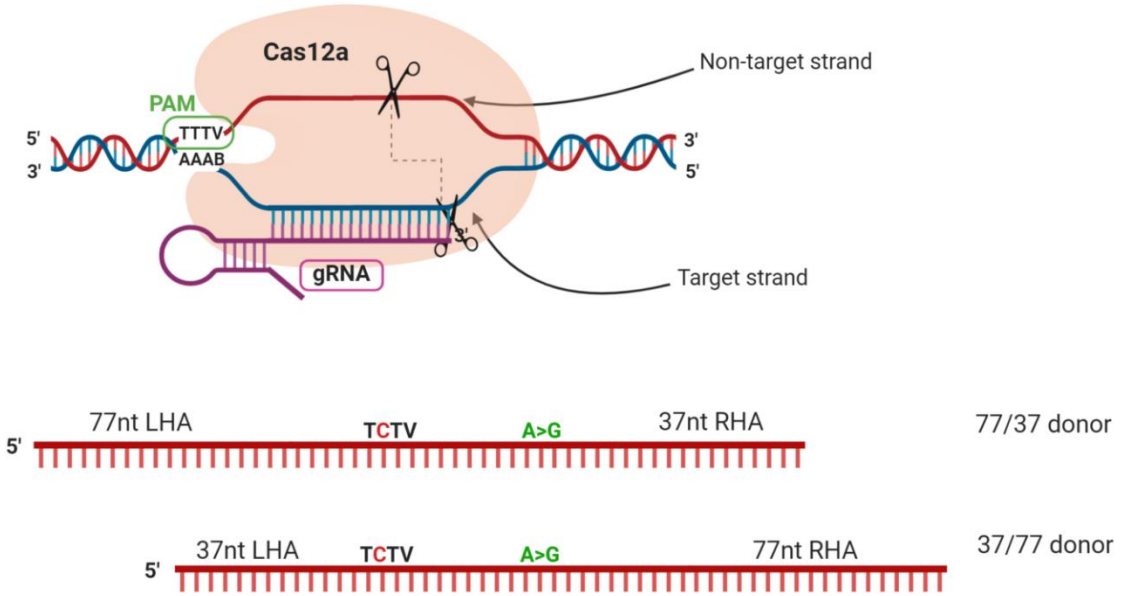


Figure 3.3: Design of the Cas12a ssODN donors. Two donors were designed, both containing the non-target strand sequence, with the A to G base correction of the W1282X mutation, and a T to C change in the PAM sequence. The donor 77/37 possessed 77nt of left homology arm, and 37nt of right homology arm. On the contrary, the 37/77 possessed 37nt of left homology arm, and 77nt of right homology arm.

Figure 3.3 - Design of the AsCas12a ssODN donors

An electroporation of AsCas12a/Guide 1 RNP using 6µM of each ssODN donors was performed in the 16HBE14o$^-$ cells containing the W1282X mutation. As controls, Cas12a protein alone without Guide 1, and donors alone were electroporated. Because electroporation resulted in many cell deaths and few cells plated in the well, the first DNA analysis was done 10 days post-transfection, once the cells reached confluency. After 3 passages (or 30 days post-transfection), a second DNA analysis was performed. The DNA was extracted, and PCRs were done using a set of primers targeting exon 23 (exon23 FW) and intron 23 (intron23 REV). The samples were sent for Sanger sequencing and were analysed using the Synthego ICE (**I**nference of **C**RISPR **E**dits) software, which, compared to TIDE software, allowed a comparison of editing efficiencies for several samples. Preliminary data for indels and knock-in efficiencies are shown in figure 3.4.

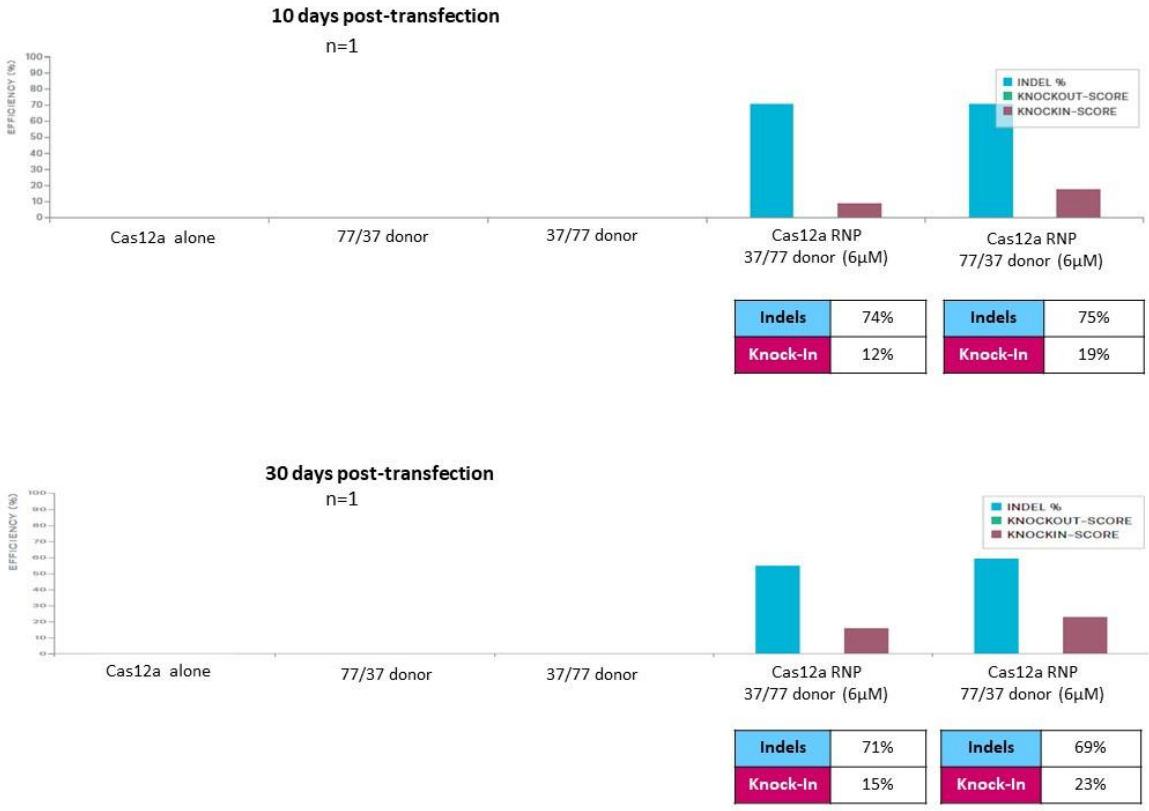


Figure 3.4: ICE analyses displaying the editing efficiencies for AsCas12a protein alone, each HDR donor template alone or for each HDR template transfected with the AsCas12a/Guide 1 RNP, at 10 days post-transfection (upper graph) and one month (3 passages) post-transfection (lower graph). The tables show for each condition of AsCas12a/Guide 1 RNP + donor, the percentage of indels (in blue) and the percentage of indels which contains the HDR template sequence (in pink). The results are shown for one experiment.
Importantly, the amount of indels corresponds to every modification observed in the sequencing file. Therefore, the percentage of indels contains the percentage of knock-in.

Figure 3.4 - Editing efficiencies at 10 days and 30 days post-transfection, comparing two different HDR donor templates for AsCas12a

From those preliminary results displayed by the ICE analyses, it can be seen, first, that all controls showed no visible indels or W1282X correction, as expected.

Second, it can be noticed that at 10 days and one-month post-transfection, a high percentage of indels of about 69% to 75% were observed. Those percentages were similar between the two ssODN donors. Comparing the HDR efficiencies for the two donors, the 77/37 donor displayed a W1282X correction of about 19%, slightly higher than the 12% observed for the 37/77 donor. This experiment has been done only once as a preliminary data to confirm that the 37/77 gRNA was indeed the most efficient as described in the paper from 2018.

Those results demonstrate that the ssODN 77/37 was the most efficient, consistent with observation in other studies (115).

Interestingly, after about 30 days (3 passages) post-transfection, it can be observed that the percentages of indels and HDR were similar to those observed at 10 days. Those results indicate that the correction tends to last with time and cell passages.

This experiment was a preliminary result for W1282X correction using Cas12a, in order to define which ssODN template was the most efficient to use for further experiments. In this experiment data, the donor 77/37 was the donor performing the best.

3.3 W1282X CORRECTION AT THE DNA LEVEL

Now that both optimal gRNA and donor were identified, the next step was to test different amounts of donors and see what effect this had, if any, on HDR efficiencies. This would enable a comparison with the Cas9 HDR experiments (performed by another PhD student in the lab), to identify which of the two techniques would be the most suitable to correct the W1282X mutation for CF. The comparisons will be described in the discussion of this chapter.

In 2018, the company IDT published on their website an experiment using Nucleofection to perform HDR. In their experiment, the authors showed high HDR efficiencies using about 3µM of oligonucleotide template (185). Therefore, the next experiment used concentrations around those 3µM as a starting point.

Three independent electroporations were performed using three different amounts of ssODN template donors (0.6µM, 2µM and 6µM) with the AsCas12a/Guide 1 RNP, as well as two controls: RNP only and 77/37 ssODN donor only. All those conditions were compared with a pcDNA empty plasmid negative control, for analysis of the editing efficiencies. Four days post-transfection, the DNA was extracted, and PCRs were performed for every experiment and condition using the primers exon23fw and intron23rev. The samples were sent for sequencing and analysed using the Synthego ICE analysis tool, which has a limit of detection of about 5% editing depending on how clear the sequencing is. In figure 3.5 is represented the average percentages of HDR efficiencies and indels formations from the three independent electroporations.

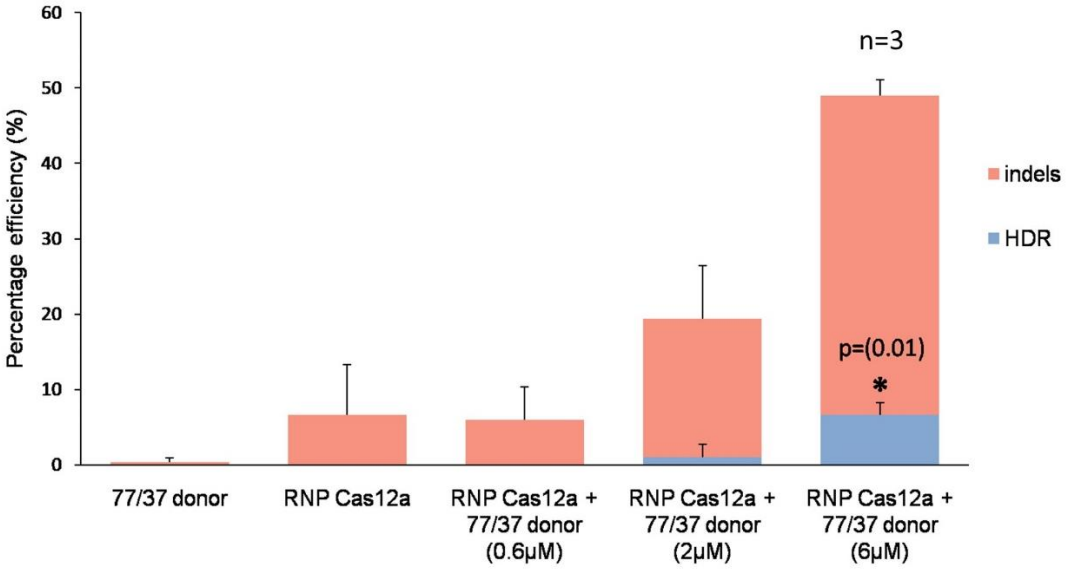


Figure 3.5: Editing efficiencies for HDR correction and indels formation at the DNA level. An electroporation was performed using pcDNA empty plasmid as a negative control, and different conditions being the 77/37 ssODN donor alone, the AsCas12a/Guide 1 RNP alone, and both the RNP and different amount of ssODN donor (0.6µM, 2µM and 6µM). Four days post-transfection, PCRs were performed around the cut site and the amplicons were sent for Sanger sequencing. The sequencing files were analysed using the ICE analysing tool to determine the average percentage of HDR (red) and indels (blue) for each condition.
The average results were performed from three independent experiments. A Ttest was performed in order to determine the significance (=p≤0.05).*

Figure 3.5 - Editing efficiencies for HDR correction and indels formation at the DNA level

From the graph in figure 3.5, it can be seen that the two controls did not display any W1282X correction through HDR, showing that both a DSB and a donor template are required to enable recombination. Noticeably, the RNP alone induced about 7% of indels, which was expected since RNPs create DSB in the DNA, which will be repaired by NHEJ and produce indels. Those 7% are also consistent with the 8.8% indels displayed by the RNP during gRNA characterisation.

Importantly, a dose-dependent increase in both HDR and indel formation, was observed as the amount of donor templates was increased. Indeed, 0.6µM of donor templates with AsCas12a/Guide 1 RNPs led to the production of about 6% indels and 0% HDR, while 2µM of donor templates with AsCas12a/Guide 1 RNPs led to the production of about 18% indels and 1% HDR. Finally, the electroporation of 6µM donor templates with AsCas12a/Guide 1 RNPs led to the production of about 42% indels and about 7% HDR.

Noticeably, when the donor template was increased by 3-fold, a ~3-fold increase in HDR and indels percentages could be observed. In addition, it can be observed than the values obtained with 6µM of donor templates (7% HDR and 42% indels) are in accordance with the values obtained in the preliminary data to determine the optimal template (section 3.2; 19% HDR and 56% indels).

It can be predicted that if the donor amount is even more increased, the efficiency of HDR could become even higher than the 7% observed. Additional electroporations could be performed using the same amount of donors plus two more increasing amounts, to test this idea. However, it could also be possible that increasing the amount of donor could further increase indel levels or induce a toxicity, which could kill the cells and decrease the HDR efficiency.

Moreover, it can be observed that adding increasing amount of ssODN donors increases the amount of indels. This was unexpected since the RNPs, which are introduced in a constant amount, are the components that are responsible for the DSB and the indels formation. An idea to explain those results could be that increasing the amount of donors might increase the chances for the donor to enter the cells and/or interact with the genomic DNA. Perhaps, following DSB, the donors might interact with the DNA ends, possibly preventing the cells from repairing the break scarlessly (Figure 3.6). The assumption is that the staggered break caused by Cas12a after cutting, would lead to ssDNA ends containing about 4 to 6nt, which are complementary to the HDR donor. It could be imagined that there would be an interaction between the free ends and the donor, preventing the repair system to join the free ends together. It can be assumed that increasing the dose of templates could increase the amount of donors entering the cells, increasing the probability for a donor to interact with the free ends. This would prevent end-repair and increase the indels frequency (Figure 3.6).

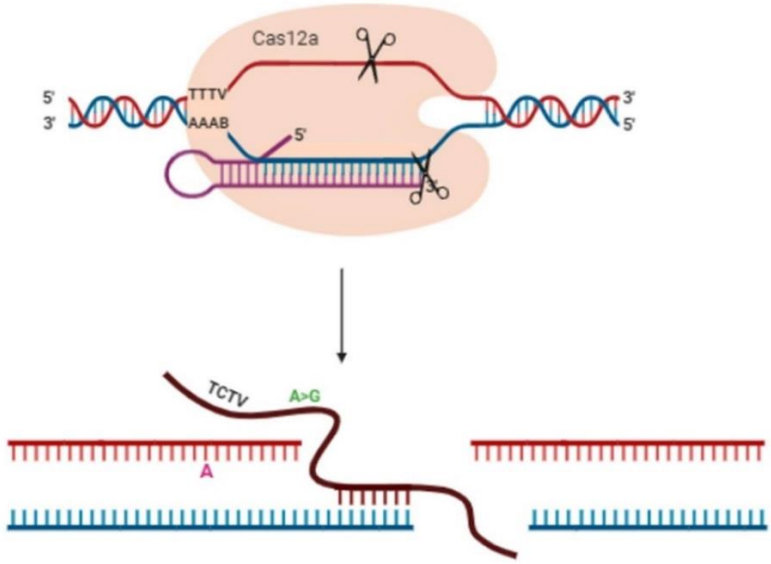


Figure 3.6: Model for the increasing indels efficiencies assumption. Following the staggered cut from Cas12a, there would be 4 to 6 nucleotides that are complementary to the ssODN donor, allowing an interaction that would prevent the repair system to join both ends perfectly.

Figure 3.6 - Model for the increasing indels efficiencies assumption

To test this hypothesis, one possibility could be to design a donor that contains mismatches at the overhang-complementary site, to assess if there is still a proportionality between the amount of donor and the percentage of indels frequency.

3.4 W1282X CORRECTION AT THE RNA LEVEL

After showing genomic DNA correction, the next step was to determine the effect of DNA editing on the RNA transcripts. The pool of transfected cells from the three previous independent experiments were used for the RNA experiments. RNA from the pool of cells was extracted from the cells and an RT-PCR was performed using a set of primer targeting the exon 21 (CFTR exon 21 FW) and the junction between exon 24 and 25 (junction 24-25 CFTR REV). The PCR products were sent for sequencing and analysed using the Synthego ICE Analysis tool (Figure 3.7).

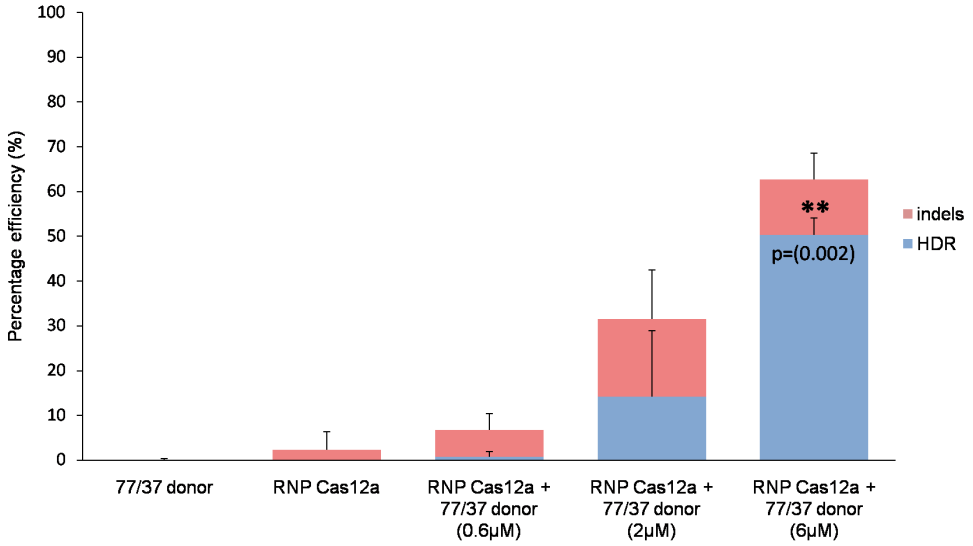


Figure 3.7: Editing efficiencies for HDR correction and indels formation at the RNA level. RT-PCR were performed to every experiments and conditions studied at the DNA level, four days post-transfection. The PCRs were performed around exon 23 and the amplicons were sent for Sanger sequencing. The sequencing files were analysed using the ICE analysing tool to determine the average percentage of HDR (red) and indels (blue) for each condition.
*The average results were performed from three independent experiments. A Ttest was performed in order to determine the significance (**=p≤0.01).*
Figure 3.7 - Editing efficiencies for HDR correction and indels formation at the RNA level

The first thing to be noticed is that, similar to the DNA results, there is an increase in the number of transcripts containing the W1282X correction, in proportion to the increase in the amount of donor template used.

Interestingly, the percentage of transcripts containing indels is almost identical to the proportion of DNA containing indels, for cells treated with AsCas12a/Guide 1 RNP + 0.6µM donor (~6% for both RNA and DNA) and with AsCas12a/Guide 1 RNP + 2µM donor (~17% for RNA vs ~18% for DNA). In contrast, there is a notable decrease in transcripts with indels, in cells treated with AsCas12a/Guide 1 RNP alone (~2% for RNA vs ~7% for DNA) or with AsCas12a/Guide 1 RNP + 6µM donor (~12% for RNA vs ~42% for DNA).

This result could be explained if the genomic DNA containing indels (showed in red in figure 3.5), gave rise to RNA containing those same indels, and those aberrant RNAs were sent for degradation through the NMD pathway. Indeed, the production of indels can, most of the time, create a premature stop codon in the mRNA sequence, by modifying the open reading frame. Those premature stop codons are recognised by the NMD pathway, which degrades the erroneous RNA.

Importantly, there is a high percentage of HDR corrected transcripts compared to the percentage of HDR corrected DNA for the same conditions. Indeed, the controls containing either AsCas12a/Guide 1 RNP alone, or 77/37 donor alone, did not show any correction at the RNA level.

However, AsCas12a/Guide 1 RNP with 0.6µM, 2µM or 6µM of 77/37 ssODN donor displayed 0.8%, 14% and 50% RNA correction respectively (in blue in figure 3.7). Those results corresponded to ~1-fold, ~14-fold and ~7-fold increase respectively, compared to the genomic DNA results from the same conditions. Those increases in percentage of corrected transcripts compared to corrected DNA could also be explained with the role of the NMD pathway. It has been previously shown that W1282X mutant 16HBE14o$^-$ cells have a lot less *CFTR* mRNA compared to the WT 16HBE14o$^-$ cells (175). Indeed, the W1282X mutation being a premature stop codon, most of the RNA containing this mutation is more likely to be degraded by NMD, shortening its half-life. Therefore, it is expected that the correction of the mutation would prevent the *CFTR* RNA from being degraded, increasing its half-life in the cells, and consequently, allowing an accumulation of corrected RNA in the cells over the mutated RNA.

3.5 W1282X CORRECTION AT THE PROTEIN LEVEL

3.5.1 - ISOLATION OF CLONAL CELL LINES

Having shown in the previous section that efficient DNA and RNA correction was achieved in the W1282X mutant 16HBE14o$^-$ cell lines, the next step was to determine if a functional corrected protein was produced. First, because it is not known whether 7% correction would be enough to see a functional correction in a population of CF cells, it was decided to make clonal populations in order to get 100% of corrected cells in the population. The purpose of analysing a 100% cell-corrected population was to confirm that the Cas12a correction led to a functional protein production. If the protein correction can be confirmed on a 100% population, it would then be feasible to attempt to determine if the 7% correction achieved in the mixed cell population was enough to give a functional benefit. To make clonal cell lines, single cells from one of the Cas12a with 77/37 ssODN donor (6µM) experiment, were plated in 96 well plates, using a FACS Aria^{TM} Fusion cell sorter. After allowing the cells to grow, 56 clones succeeded to make a cell population. The DNA was extracted, and PCRs were done using the set of primers targeting exon 23 and intron 23.

The PCR products were sent for Sanger sequencing and 42 sequencing profiles yielded data suitable for analysis (Table 3.1).

	Clonal populations	PAM W1282X
W1282X/W1282X	6/42	
WT/indels	6/42	
WT/WT	3/42	
WT/W1282X	1/42	
Indels/indels	13/42	
W1282X/indels	13/42	

Table 3.1: Table showing the number of homozygous corrected (the orange row), homozygous non-edited, heterozygous or indels-containing clonal populations. The cells from the condition AsCas12a:Guide 1 RNA + 37/77 donor (6µM), from the experiment 3, were single cell sorted to form clonal populations. Forty-two clonal populations were analysed, and the amount of populations containing two W1282X alleles (non-corrected), one WT + one indel-containing alleles, two WT alleles (homozygous corrected), one WT + one W1282X alleles (heterozygous corrected), two indels-containing alleles, and one W1282X + one indel-containing alleles, were reported in the table above. A representative chromatogram of the sequencing for each category is shown on the right column. The yellow rectangle shows the site of the nucleotide change in the PAM and the red rectangle shows the site of the W1282X mutation.

Table 3.1 - Number of homozygous W1282X-corrected, homozygous non-edited. heterozygous or indels-containing clonal populations from AsCas12 HDR experiment

Amongst those 42 analysed cell lines, 6/42 (~14%) cell lines were not modified, 3/42 (~7%) were homozygous for the W1282X correction, and the remaining were heterozygous mostly with indels (Table 3.1). Knowing that each cell is biallelic, there was therefore 84 alleles analysed in this experiment. Amongst those 84 alleles, 26/84 (~31%) of the cells were W1282X and thus, not modified. Moreover, 45/84 (~53%) were indels-containing alleles, and 13/84 (~15%) had the W1282X mutation corrected by HDR (Table 3.1).

Interestingly, the percentages given by the ICE analysis in figure 3.5 were ~50% of the genomic DNA that was not modified, ~42% that was indels-containing, and ~7% that was corrected by HDR. The differences from these percentages and the ones given by the analysis of the clonal populations could be due to probabilities and the smaller amount of clones analysed (42 clones) compared to the amount of genomic DNA analysed by sequencing.

3.5.2 - PROTEIN TRANSLATION

The corrected cell lines being produced, it was then possible to assess the protein translation and function by Western Blot and Ussing Chamber respectively.

First, to check for protein translation, the heterozygous and homozygous clonal populations were cultured into a liquid–liquid culture interface in snapwells^{TM}. As controls, W1282X mutant 16HBE14o$^-$ and WT 16HBE14o$^-$ cells were also cultured in the liquid-liquid interface. After 7 and 8 days in culture, the **T**rans**E**pithelial **E**lectrical **R**esistance (TEER) was measured using an epithelial Volt/Ohm meter. Preliminary results for this experiment are shown in Table 3.2.

Cells	Cell information	TEER
16HBE14o$^-$ WT	Positive control expressing CFTR	1227 Ω.cm²
16HBE14o$^-$ W1282X	Negative control not expressing CFTR	1439 Ω.cm²
16HBE14o$^-$ W1282X clone 1	Heterozygous corrected	200 Ω.cm²
16HBE14o$^-$ W1282X clone 2	Homozygous corrected	156 Ω.cm²
16HBE14o$^-$ W1282X clone 5	Heterozygous corrected	1160 Ω.cm²
16HBE14o$^-$ W1282X clone 17	Heterozygous corrected	363 Ω.cm²
16HBE14o$^-$ W1282X clone 19	Heterozygous corrected	790 Ω.cm²
16HBE14o$^-$ W1282X clone 24	Heterozygous corrected	1250 Ω.cm²
16HBE14o$^-$ W1282X clone 32	Heterozygous corrected	768 Ω.cm²
16HBE14o$^-$ W1282X clone 36	Homozygous corrected	140 Ω.cm²
16HBE14o$^-$ W1282X clone 47	Homozygous corrected	134 Ω.cm²

Table 3.2: TEER measurement results for each clonal population, as well as for the 16HBE14o$^-$ WT positive control and the 16HBE14o$^-$ W1282X negative control. All the homozygous and heterozygous corrected clonal populations were cultured in liquid-liquid interface to form tight junctions. After seven and eight days in culture, TEERs were measured using an EVOM (section 2.7.2).
In 2007 a paper described for NHBE cells, a peak of TEER value of about 766+/-154 Ω.cm^2 on the 8th day (186).
In yellow were highlighted the values of resistance considered too low for the epithelium to form tight junctions.
Those measurements are the results of a unique experiment.

Table 3.2 - TEER measurement for each homozygous and heterozygous W1282X-corrected clonal population

A western blot of each clonal population from the table 3.2, cultured on liquid-liquid interface, was performed, to confirm the presence of protein. The W1282X 16HBE14o$^-$ cells were not expected to express CFTR proteins due to unstable mRNA. Indeed, the presence of a premature stop codon in the transcript serves as signal to send the mRNA for degradation through the Non-sense Mediated-Decay (NMD). However, if the *CFTR* gene is corrected, CFTR protein is expected to be expressed.

The cells were extracted from the snapwells^{TM} and the clones were sent to a collaborator in Lisbon for western blot experiments (Figure 3.8).

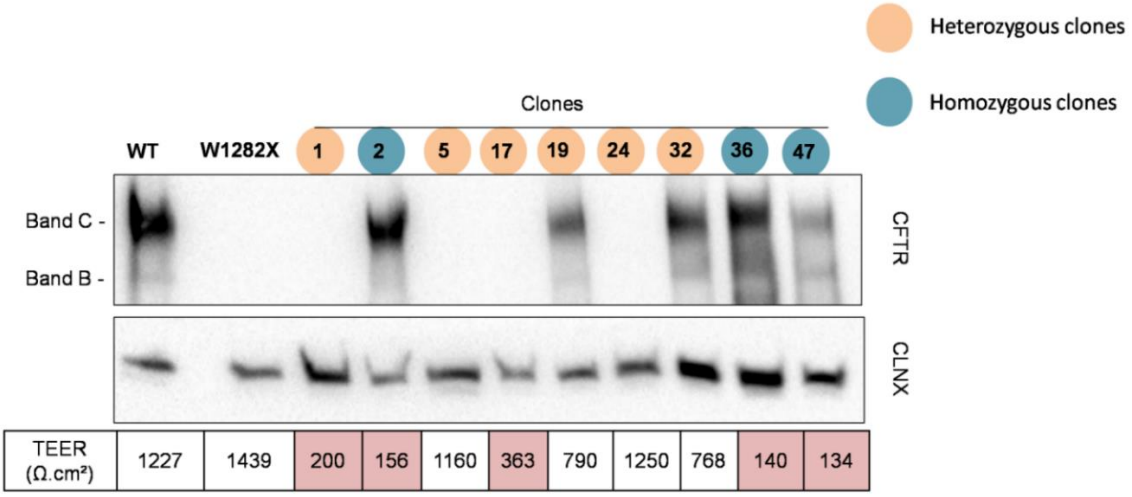


Figure 3.8: Western blot analysis of the 16HBE14o$^-$ W1282X corrected clones. All the samples grown on liquid-liquid interface and for which the TEER measurements were done, were sent to Lisbon for western blot analysis.
The samples were run in a 4-7% polyacrylamide gel. The proteins were transferred to a PVDF membrane that was blocked with 5% of non-fat milk in TBST. CFTR (UNC596) and Calnexin (CLNX) antibodies were used at a 1.3000 dilution. The Calnexin was used as a loading control. The TEER values for each clonal population are shown in the table under the western blot.
This result is the result of a unique experiment.
Western Blot made in collaboration with Lucia Santos from the University of Lisbon

Figure 3.8 - Western blot analysis of the W1282X-corrected clones.

First, the western blot results show expression of the CFTR protein for the WT 16HBE14o$^-$ control cells and no CFTR protein expression for the mutant 16HBE14o$^-$ W1282X controls, as expected. Importantly, the western blot shows that every homozygous cell, as well as 2/6 heterozygous clones were expressing the CFTR protein. The results for the heterozygous clones were unexpected.

However, during the course of this thesis, it was discovered that the 16HBE14o$^-$ cells have an insertion of an unknown size SV40 DNA, in the intron 6 of one of the *CFTR* allele. This insertion appears to have arisen during the immortalisation process and leads to an aberrant *CFTR* mRNA from that allele. Therefore, only one *CFTR* allele in the 16HBE14o$^-$ cells is functional (Figure 3.9; (175)).

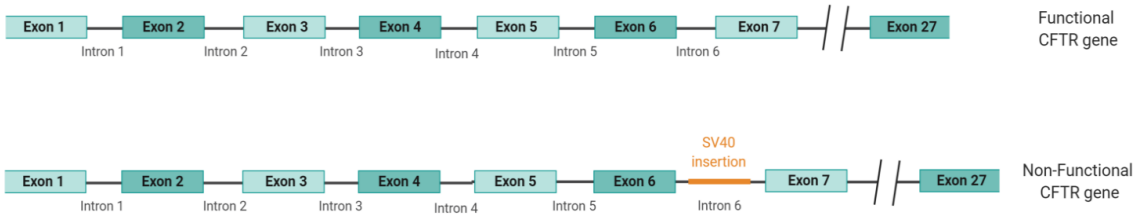


Figure 3.9: Graphic representation of the SV40 insertion into the intron 6 of the CFTR gene in 16HBE14o$^-$ cells. The insertion leads to a non-functional allele.

Figure 3.9 - Graphic representation of the SV40 insertion into the intron 6 of the *CFTR* gene in 16HBE14o$^-$ cells

The western blot analysis could now be explained by the correction of either the functional or the non-functional *CFTR* allele. Indeed, the expression of the CFTR protein suggests that the correction happened in the functional allele. In the same way, if the cells do not express CFTR protein, it would indicate that the correction happened on the non-functional allele.
To conclude, this result indicates that correcting *CFTR* mutations by HDR using Cas12a can lead to the translation of CFTR protein. Moreover, the western blot also shows that clones 19 and 32 were clones containing the W1282X correction in the functional allele.

3.6 DISCUSSION AND COMPARISON BETWEEN CAS12A AND CAS9

In this chapter, it was shown that Cas12a was able to efficiently correct the W1282X mutation in the *CFTR* gene, and this correction led to the production of full length CFTR protein. However, further studies would be required to confirm that the produced CFTR protein is functional.

3.6.1 - TEER RESULTS ANALYSIS

In table 3.2, it is important to notice that only about half of the cells achieved high transepithelial resistance. The other half had a resistance too low to be able to do any functional assay on them by Ussing Chamber. Noticeably, all the homozygous clones and about half of the heterozygous clones did not attain a suitable resistance.

One theory to consider could be that the correction of the functional allele, somehow, would have a negative impact on the resistance. Following this theory, one would predict that the heterozygous clones with low resistance could be the ones with the correction on the functional allele, while the homozygous clones, which for sure will have a correction on the functional allele, would definitely have a low resistance (Figure 3.10).

However, the reason why the correction of the functionally expressed allele would decrease the resistance would be unknown.

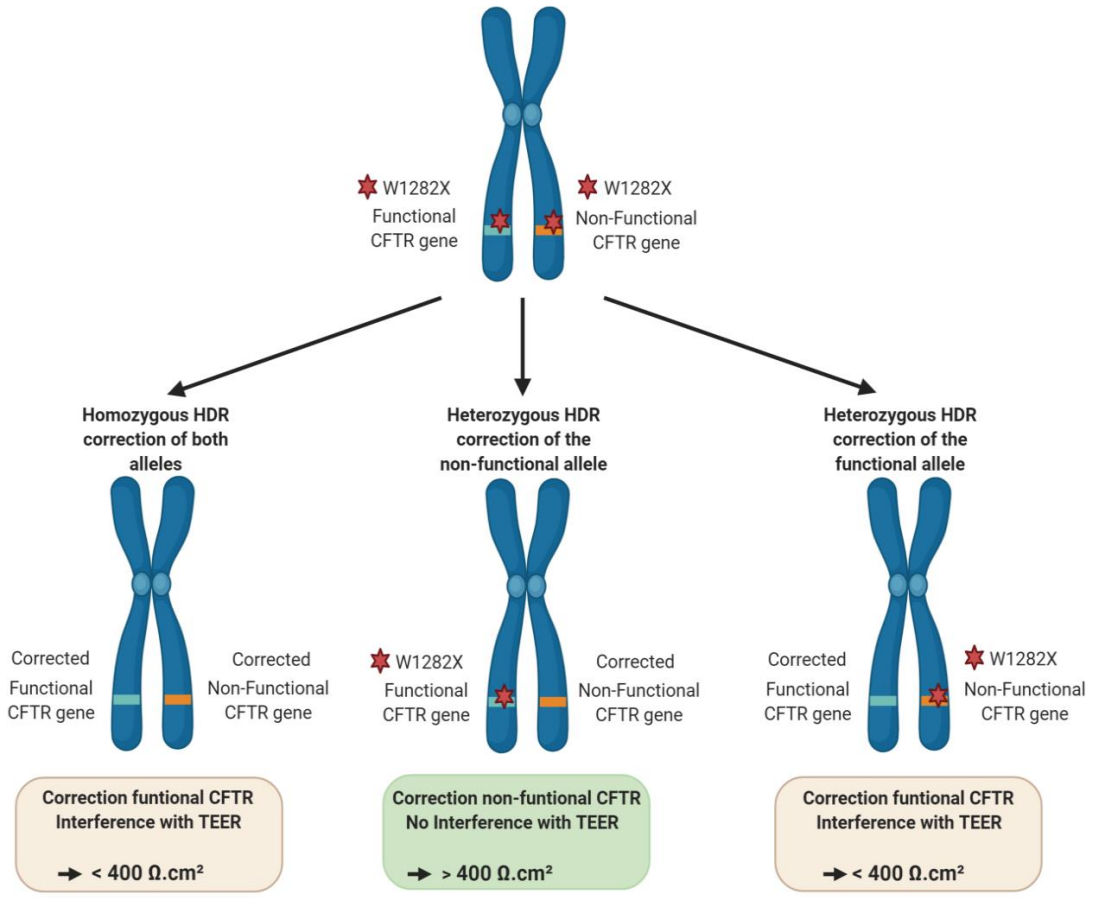


Figure 3.10: Assumption that correction of the functional allele somehow interferes with the trans-epithelial electrical resistance of the epithelium.

Figure 3.10 - Model explaining the assumption that correction of the functional allele interferes with the TEER of the epithelium

Noticeably, the western blot result showed that the clones that were corrected in the functional allele did not match the clones that had a low resistance (Figure 3.8). Therefore, the hypothesis that the CFTR correction would prevent the cells from forming tight junction is refuted.

Another possibility to explain the low resistance is that the inability to attain a suitable resistance could be a result of a stress due to the single-cell cloning. Indeed, 16HBE14o$^-$ cells growing at high density in normal culture conditions, may behave differently when grown as single cell clones. However, if it was the case, it would be expected that every clone would have a low resistance, since they all had the same treatment.

Nevertheless, it is still possible that those modifications may have been caused randomly on some cells, explaining why only half of them do not have resistance. It is important to note that both control cells have not been single cell cloned, which could be why they have high resistance. This hypothesis could be tested by taking a control cell line such as WT 16HBE14o$^-$ and making single cell cloning both on a coated or non-coated plate. After growth, the cells could be cultured in liquid-liquid interface, and the ratio of resistant over non-resistant cell lines for both conditions could be analysed. If half the cells plated on non-coated cells are still non-resistant, but the cells plated on coated plate have a higher ratio of resistant cells, it could indicate that the single cell sorting was the cause of the resistance loss.

The third theory could be that maybe those low resistance cells did not have enough time to form a tight epithelium and maybe needed more time to grow. To confirm this hypothesis, the cultures need to be done again and the resistance need to be measured over time. If only the cells in yellow in the table are the ones that keep not having a resistance, it would mean that it is those cells specifically that lost their ability to form tight junctions, then, bringing back to the second theory.

3.6.2 - CFTR FUNCTIONAL ASSAY

To test CFTR protein function, Ussing chamber experiments should be done on the liquid-liquid cultured cell lines that are resistant. During this PhD, I had the occasion to go to London for two weeks to learn Ussing Chamber experiments with our **S**trategic **R**esearch **C**entre (SRC) collaborators. Moreover, I had the opportunity to attend the "Hands-on Workshop Epithelial Systems: Physiology and Pathophysiology" in Lisbon in 2017, where I learnt different techniques to assess *CFTR* translation and function.

Finally, Maximillian Woodall from our SRC in London, kindly came to our lab for one week, to help us develop the Ussing chamber technique in the University College Cork (UCC). This Ussing chamber set up was also kindly assisted by Dr. Niall Hyland from the APC department of UCC.

Using all that information, the Ussing chamber experiments will be able to be done in the near future.

3.6.3 - COMPARISON BETWEEN CAS12A AND CAS9

Cas12a was shown to correct *CFTR* efficiently, with the appropriate gRNA and ssODN donor template. It was shown in this chapter that this enzyme could be used to correct mutations in the *CFTR* gene. However, is it the best technique to use?

In parallel to the experiments done with Cas12a, Lucia Santos, from the University of Lisbon, came to Dr. Patrick Harrison's Lab as part of her PhD, for 18 months. During this period, Lucia did the Cas9 experiments using HDR, to correct W1282X mutation. The results that she found were around 18% W1282X DNA correction followed by around 60% of corrected transcripts. Those results were obtained using 2µM of ssODN donor.

Cas12a results showing an efficiency of 7% DNA correction and 50% correction of the transcripts, using 6µM of donor, those results tend to indicate that Cas9 system seems to be more efficient. Indeed, on top of having a higher efficiency of correction, the amount of donor used to reach the optimum correction was lower for SpCas9. To bring the technique to therapeutics, it is important to use as little exogenous DNA donor template as possible to avoid any toxicity. The enzymes being ordered from IDT company, the electroporations were done following IDT protocols (section 2.3.3b). The comparisons of the amounts of components used for both enzymes are shown in table 3.3.

	Cas9	Cas12a
protein	18pmol (1.5µM)	63pmol (5.25µM)
guide	1.8µM	6.25µM
donor	0.6µM; 2µM; 6µM	0.6µM; 2µM; 6µM

Table 3.3: Comparison between the conditions used for Cas9 and for Cas12a experiments, following IDT protocol for Neon electroporation.

Table 3.3 - Comparison between the conditions used for Cas9 and for Cas12a experiments

Noticeably, the amount of Cas enzyme and gRNA electroporated were 3.5-fold higher for Cas12a conditions compared to Cas9 conditions. It can be concluded that in addition of Cas9 being more efficient with a lower amount of donor, it was also more efficient using less gRNA and Cas enzyme.

Those results show that SpCas9 seemed to be a more efficient and safer technique for HDR-based correction of *CFTR* mutations. Indeed, using the lowest doses as possible for therapies is a requirement to avoid possible negative side-effects.

However, to have a proper comparison between the two enzymes, the same amounts of reagents need to be used. A paper from 2019, used the same conditions of electroporation for both enzymes and showed a higher HDR efficiency using Cas12a over Cas9. The enzymes were ordered from the Aldevron company, while the enzymes used in this chapter were ordered from IDT. Therefore, it could be interesting to test this Aldevron Cas12a with both the Aldevron and IDT Cas9, to check if a higher HDR efficiency for Cas12a could be attained (119). Moreover, as described in section 1.8.2, IDT also made available a new Cas12a (Alt-R A.s.Cas12a (Cpf1) *Ultra* Nuclease), which displayed an increased editing efficiency compared to the AsCas12a V3 used in this chapter.

To explain the lower efficiency of Cas12a over Cas9, one assumption, described in the section 1.8.2f, could be the trans-cleavage activity of Cas12a. Indeed, Cas12a possesses the capacity to cut ssDNA when the enzyme is activated. Because the ssODN donor template is a ssDNA, it could be possible that Cas12a cleaves the template, reducing the HDR efficiency. To test this, it could be possible to make a ssODN protected with phosphorothioates (PTO) modifications, which prevent Cas12a from cutting. If a direct comparison between the previously used ssODN and a PTO protected ssODN shows a higher HDR efficiency for PTO ssODN, then the hypothesis for trans-cleavage of the donor would be confirmed.

Concerning the safety for both Cas9 and Cas12a, the indels observed for Cas12a were about 40% when the cells were electroporated with the AsCas12a/ Guide 1 + 6µM of 37/77 ssODN donor template. For Cas9, the amount of indels found, when the cells were electroporated with Cas9/gRNA RNP + 2µM of ssODN donors, was around 20%, about 2-fold lower for a greater HDR efficiency. Because indels can irreversibly damage the *CFTR* gene, this result makes Cas9 a safer technique to use.

However, another harmful effect of genome editing is the potential off-target cleavages created by the nucleases. The off-target effects have not yet been analysed for each enzyme in this experiment. In the introduction, it was described that Cas12a has been shown to be more specific with less off-target effects than SpCas9. Therefore, off-target analyses would need to be done for each enzyme, in order to compare the off-target profiles. Many techniques for off-target analyses have been described such as the T7E1 assay, Deep sequencing, GUIDE-seq or CIRCLE-seq (108).

Although the HDR technique was shown here to work efficiently, the amount of indels produced, even for SpCas9, is too high and might make it challenging to develop this technique into clinic. To overcome this downside of HDR, another technique called “Base-editing” was analysed for W1282X correction.

CHAPTER 4 - PRECISE GENOME EDITING: CORRECTING A SINGLE SPECIFIC MUTATION BY BASE EDITING

In 2016, David Liu's group in Harvard (USA) and Akihiko Kondo's group in Kobe University (Japan), developed two different cytidine base editors that can convert cytidine nucleotides into thymine, at a specific location in the genome (155, 187). One year later, David Liu's lab published a paper describing a new base editor called **A**denine **B**ase **E**ditor (ABE), which can convert adenine bases into guanine. Because the base editing technique uses a mutant nickase Cas9, that cuts only one strand of DNA, it appeared to be safer that the previously described Cas9 HDR technique, potentially limiting the number of indels and genomic rearrangements (section 1.11; (155)).

Another critical point is that this ABE could potentially change any stop codon mutation (TGA, TAA or TAG) into a Tryptophan (TGG) or convert TGA into Arginine (CGA) and TAA or TAG into Glutamine (CAA, CAG). There are three different outcomes that can emerge from the stop codon base editing: first, it is possible that the corrected base restores the codon of origin, leading to the original gene sequence. Second, it is possible that the codon correction creates a missense mutation, where the original codon was a different amino acid than the one resulting from the base editing, but the change would not significantly affect the function of the encoded protein. Third, the missense mutation leads to a disease-causing codon (see section 1.11; (158)).

The W1282X mutation (5'-TGG-3' → 5'-TGA-3') is a G>A mutation (c.3846G>A) that creates a premature stop codon. Theoretically, an ABE could correct the A mutation back to a G, restoring the codon to its original sequence. However, using the ABE 7.10 developed in 2017, the correction of W1282X was not possible because the mutation was outside the window of editing of the technique (Figure 4.1A). Cell transfection on HEK cells using this plasmid were done by a colleague in the lab, however, the results were negative as predicted (data not shown).

However, since then, many engineered SpCas9 variants have been described, recognising different variations of the 5'-NGG-3' PAM. SpCas9-NG has been designed at the end of 2018, and has been shown to recognise 5'-NGD-3' PAM (D being a A, T or G) (166). At the same time, David Liu's group created a more efficient version of ABE called ABEmax (161). A few months later, XingXu Huang's group in Beijing (China) fused the ABEmax with the SpCas9-NG to make ABEmax-NG (188). With this new base editor, it was theoretically possible to correct the W1282X mutation (Figure 4.1B).

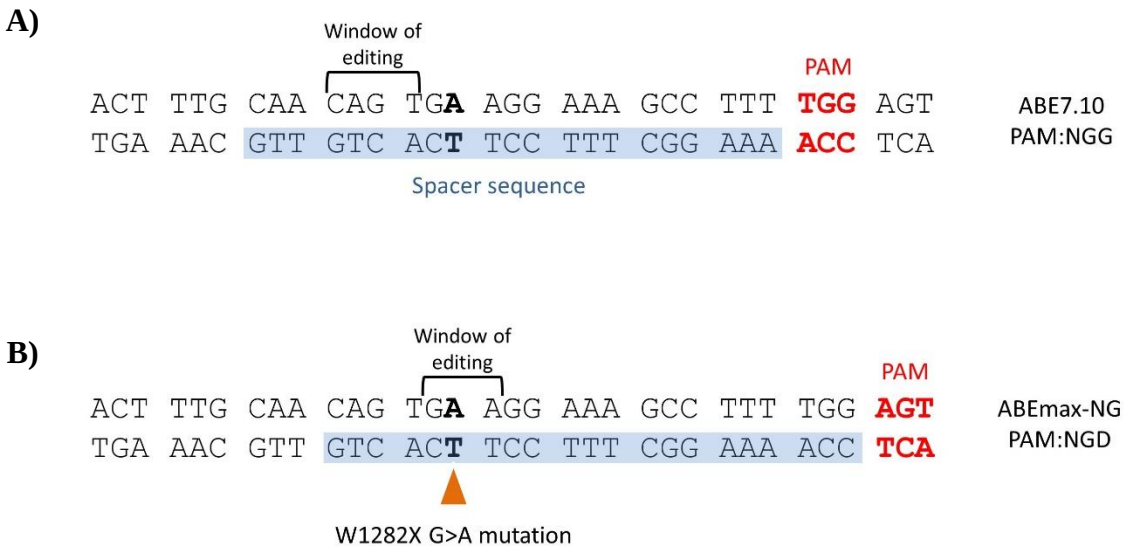


Figure 4.1: PAM positions and windows of editing for (A) ABE 7.10 and for (B) ABEmax-NG. With the ABE 7.10, the mutation W1282X was out of the correction window. With the ABEmax-NG, the mutation was inside the editing window.

Figure 4.1 - PAM positions and windows of editing for ABE 7.10 and ABEmax-NG

Importantly, one thing to notice from the figure 4.1B, is that, inside the editing window is located another adenine. Therefore, this adenine could also potentially be edited into a guanine. If that happens, the next codon (5'-AGG-3') will be changed from an arginine to a glycine. Looking at the CFTR2 database (44), where most of the CF-causing mutations in *CFTR* are referenced, a mutation was described at this site. The c.3848G>T mutation changed the original arginine (AGG) into a methionine (ATG). This mutation was considered as a CF-causing mutation (44). However, no CF-causing mutations were described for an arginine (AGG) changed into a Glycine (GGG).

Looking at the Cystic Fibrosis Mutation Database (CFTR1) (42), where every mutation found on CF patients are recorded, one c.3847A>G mutation corresponding to a change from arginine to glycine, has been discovered. This mutation has been observed in only one patient, who also had the common F508del mutation (42). The Arginine at this position is a highly conserved amino acid, and the mutation has been predicted by Polyphen as probably damaging.

Moreover, the programme Align-GVGD classified this mutation as likely to interfere with CFTR protein function. PolyPhen (Polymorphism Phenotyping) is a tool predicting possible impact of an amino acid substitution on the structure and function of a human protein. Align-GVGD is a tool using sequence alignments to predict if missense substitutions could be neutral or deleterious. This will need to be tested experimentally.

4.1 W1282X BASE EDITING AT THE DNA LEVEL

The hypothesis was to see if the W1282X mutation could be corrected using ABEmax-NG with a suitable gRNA. Because 16HBE14o$^-$ cells were very difficult to transfect with plasmids, HEK FLP W1282X EMG in21-in24 cells were used (see section 2.1). Those HEK cells were modified to have a *CFTR* transgene allele under the control of an active promoter (176), since the endogenous *CFTR* promoter was not active in HEK293 cells (Figure 4.3). The cells were transfected with either pcDNA empty plasmid, ABEmax-NG alone, gRNA A_6 alone recognising the 5'-AGT-3' PAM (Figure 4.2A), gRNA A_7 alone recognising the PAM 5'-GAG-3' (Figure 4.2B), ABEmax-NG + gRNA A_6 plasmids and ABEmax-NG + gRNA A_7 plasmids as a negative control. The transfections were done using Lipofectamine 3000, which showed good efficiency with low cell death in HEK cells.

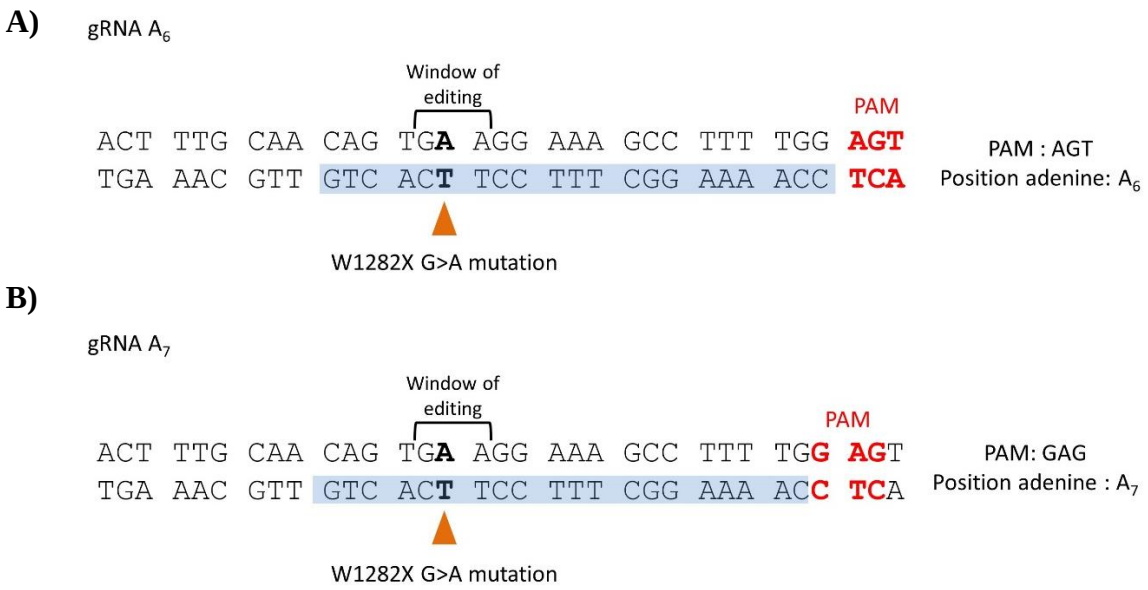


Figure 4.2: Spacer sequences, localisation and windows of editing for (A) gRNA A_6, which has the PAM 5'-AGT-3' and B) gRNA A_7, which has the PAM 5'-GAG-3'.

Figure 4.2 - Spacer sequences, localisation and windows of editing for gRNA A_6 and A_7

Four days post-transfection, the DNA was extracted, and a PCR was done using a set of primers targeting the junction between exon 22 and intron 22 (ex22-in22 FW) and the beginning of intron 23 (in23 REV EMG). Theoretically, if the primers bind the endogenous *CFTR* gene, the PCR product would be 15kb long, which is too big to be amplified. However, if the primers bind the *CFTR* transgene, the amplicon size would be 915bp. Therefore, only this PCR product would be amplified (Figure 4.3).

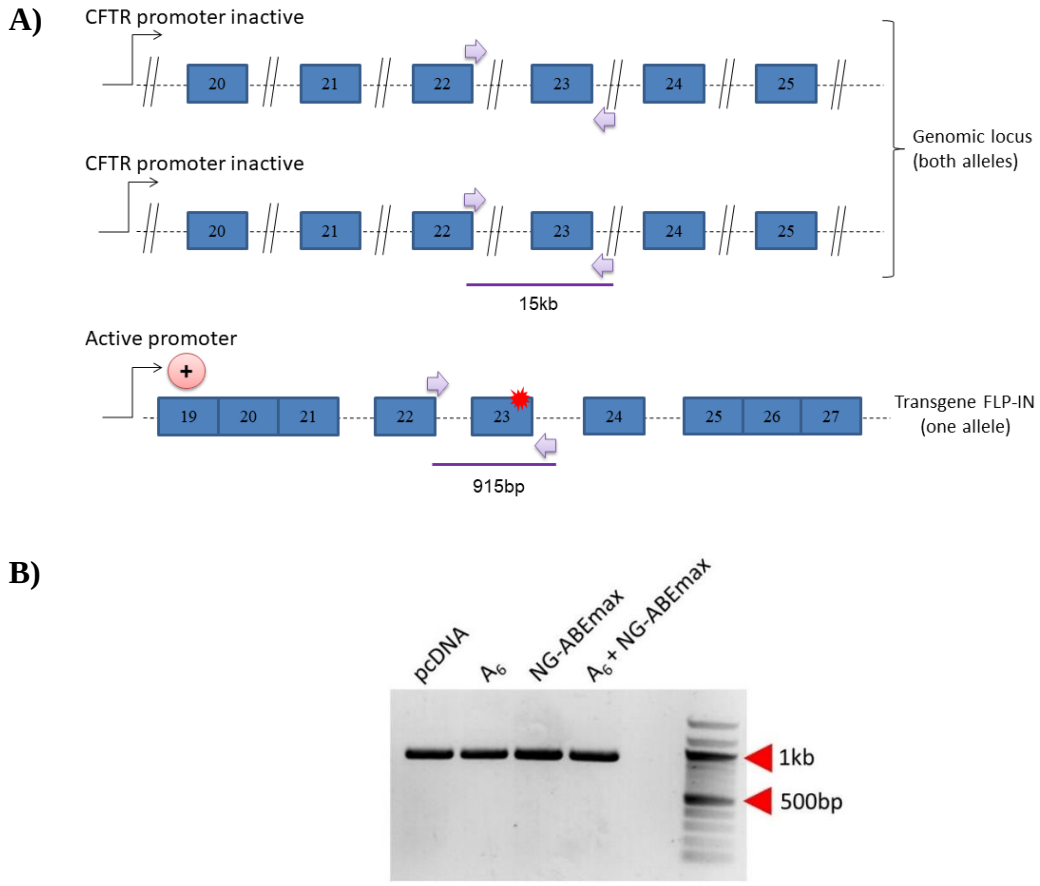
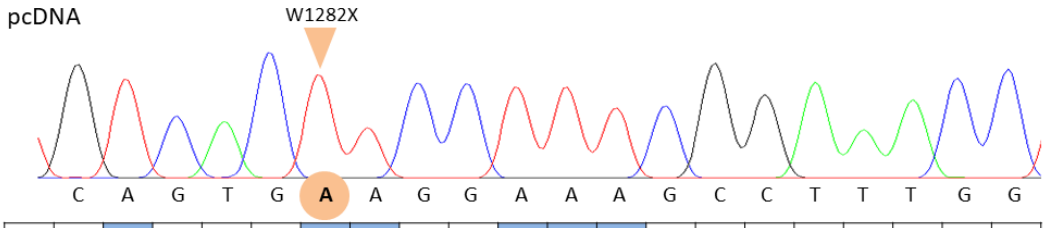


Figure 4.3: A) Schematic representation of the CFTR genotype in HEK Flp-In 293 EMG i21-i24 W1282X cells. The endogenous CFTR gene is present but not expressed, while a cDNA of the CFTR gene containing shorter introns 21 to 24, has been integrated into an active region of the genome. This cDNA contained the mutation W1282X. The arrows show the localisation of the primers binding sites. B) PCR results of the genomic DNA from transfected cells, using primers ex22-in22 FW and in23REV. HEK Flp-In 293 EMG i21-i24 W1282X cells were transfected using Lipofectamine 3000. The negative control was a transfection with an empty plasmid pcDNA, and the working conditions were gRNA A_6 alone, gRNA A_7 alone (data not shown), NGE-ABEmax alone and both gRNA A_6 + NG-ABEmax and gRNA A_7 + NG-ABEmax (data not shown). Four days post-transfection, the DNA was extracted, and PCRs were performed around the W1282X mutation. The expected size of the amplicons is 915bp.

Figure 4.3 - Schematic representation of the *CFTR* genotype in HEK Flp-In 293 EMG i21-i24 W1282X cells, and PCR results of base editing at the DNA level.

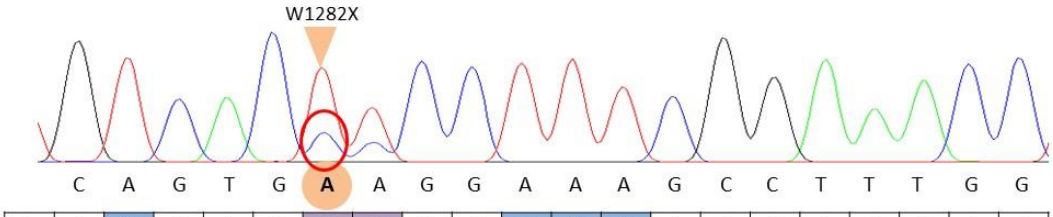
The PCR products were purified and sent for Sanger sequencing. The sequencing results were then analysed by EditR to determine the percentage of correction (Figure 4.4). EditR is a software using an algorithm to estimate the percentage of base editing from a Sanger sequencing. The results are shown as a table giving the percentage of each base in the chromatogram.

pcDNA



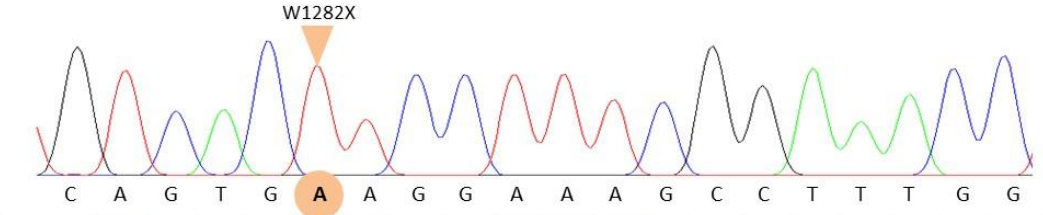
A	0	100	0	0	0	100	100	0	0	100	100	100	0	0	0	0	0	0	0	0
C	100	0	0	0	0	0	0	0	0	0	0	0	0	100	100	0	0	0	0	0
G	0	0	100	0	100	0	0	100	100	0	0	0	100	0	0	0	0	0	100	100
T	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100	100	100	0	0

RNP NG-ABEmax + gRNA A_6



A	0	100	0	0	0	76	74	0	0	100	100	100	0	0	0	0	0	0	0	0
C	100	0	0	0	0	0	0	0	0	0	0	0	0	100	100	0	0	0	0	0
G	0	0	100	0	100	24	26	100	100	0	0	0	100	0	0	0	0	0	100	100
T	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100	100	100	0	0

RNP NG-ABEmax + gRNA A_7



A	0	100	0	0	0	100	100	0	0	100	100	100	0	0	0	0	0	0	0	0
C	100	0	0	0	0	0	0	0	0	0	0	0	0	100	100	0	0	0	0	0
G	0	0	100	0	100	0	0	100	100	0	0	0	100	0	0	0	0	0	100	100
T	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100	100	100	0	0

Figure 4.4: Evidence of efficient W1282X base correction at the DNA level. The amplicons obtained in figure 4.3B were purified and sent for Sanger sequencing. The chromatograms were analysed through the EditR tool. In this figure are shown the pcDNA, NG-ABEmax + gRNA A_6 and NG-ABEmax with gRNA A_7 samples. The gRNA A6 alone, gRNA A7 alone and NG-ABEmax alone analyses displayed results similar to the pcDNA negative control (data not shown). The orange circles show the localisation of the target adenine responsible for the W1282X mutation. The red circle shows the correction from A>G of the W1282X mutation.

This result is the result of a unique experiment (n=1).

Figure 4.4 - Evidence of efficient W1282X base correction at the DNA level

In figure 4.4 are shown three samples; pcDNA, ABEmax-NG + gRNA A_6 and ABEmax-NG + gRNA A_7. The other samples: gRNA A6 alone, gRNA A7 alone and ABEmax-NG alone all had the same profiles as pcDNA (data not shown).

Interestingly, the Sanger sequencing from all the samples had very low background, showing no detectable indels and making the base editing analysis more precise. Looking at the pcDNA control sample, a very neat profile containing 100% of 5'-TGA-3' sequence, which corresponds to the STOP codon, is observed. This result confirmed that the primers did not amplify the endogenous *CFTR* gene, which contains a 5'-TGG-3' sequence. However, when ABEmax-NG was transfected with the gRNA A_6, two new peaks were clearly seen. Those peaks corresponded to the conversion of adenine into guanine. The adenine circled in red matched the one responsible for the W1282X mutation. This adenine was corrected at 24%. However, the next adenine was also edited at 26%.

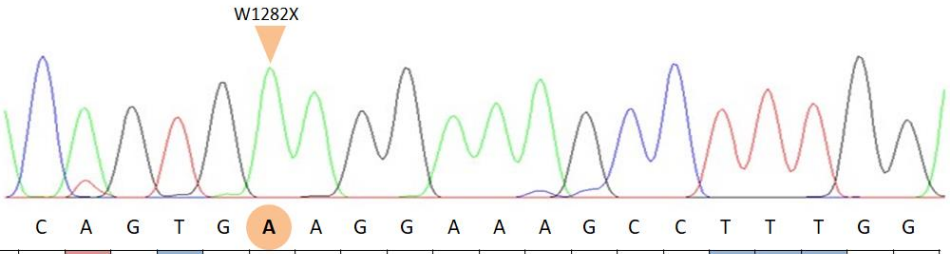
Those results demonstrated that the W1282X mutation was successfully edited, which was the first time this mutation was ever base edited. The results also showed that the next adenine was edited, changing the next codon from an arginine into a glycine.

As explained in section 4, the Cystic Fibrosis Mutation Database (CFTR1) indicated a possible damaging role of this mutation for CFTR protein function (42). However, this mutation has never been proved to have a negative effect, and it would be interesting to study the effect of changing the arginine at position 1283 into a glycine to determine if that mutation could be CF-causing. In contrast, when ABEmax-NG + gRNA A_7 plasmids were transfected in the cells, no editing was observed. This result was expected since the PAM for this gRNA, which was a 5'-GAG-3' PAM, was different from the 5'-NGD-3' PAM recognised by the Cas9 enzyme from the ABEmax-NG plasmid.

4.2 W1282X CORRECTION AT THE RNA LEVEL

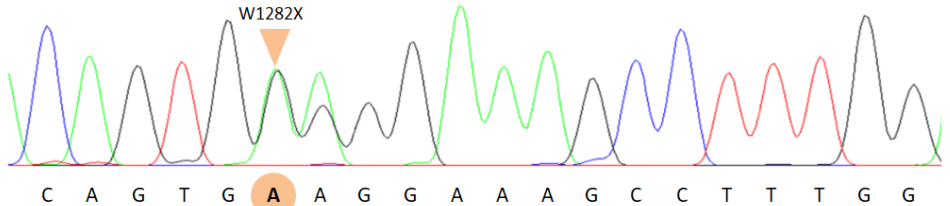
Because the W1282X-containing mRNA in cells is unstable due to NMD, with a reduction of the quantity of *CFTR* mRNA compared to WT *CFTR* mRNA (189), the next hypothesis to test was that an increase in corrected transcripts should be observable in the cDNA sequences obtained from edited cells. This would be due to a stabilisation of the corrected mRNAs. To test this hypothesis, the RNA was extracted in cells taken 4 days post-transfection and an RT-PCR was performed using a set of primers targeting the exon 21 (CFTR exon 21 FW) and the junction between exon 24 and 25 (junction 24-25 CFTR REV). The PCR samples were sent for Sanger sequencing. The sequencing data were analysed using EditR (Figure 4.5).

pcDNA



	C	A	G	T	G	A	A	G	G	A	A	A	G	C	C	T	T	T	G	G
T	0	16	0	100	0	0	0	0	0	0	0	0	0	0	0	100	100	100	0	0
G	0	0	100	0	97	0	2	100	100	0	0	0	100	0	0	0	0	0	100	100
C	100	0	0	0	0	0	0	0	0	0	0	5	0	100	100	0	0	0	0	0
A	0	84	0	0	3	100	98	0	0	100	100	94	0	0	0	0	0	0	0	0

ABEmax-NG + gRNA A_6



	C	A	G	T	G	A	A	G	G	A	A	A	G	C	C	T	T	T	G	G
T	3	3	0	96	0	0	0	0	0	0	0	0	0	0	0	100	99	100	0	0
G	0	0	100	4	100	50	39	100	98	0	0	0	100	0	0	0	1	0	100	99
C	97	0	0	0	0	0	0	0	0	0	0	2	0	100	100	0	0	0	0	0
A	0	97	0	0	3	50	60	0	2	100	100	98	0	0	0	0	0	0	0	1

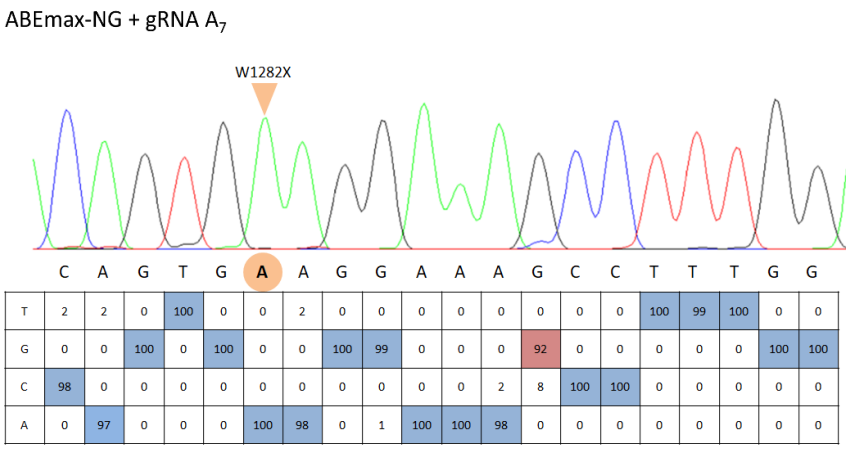


Figure 4.5: Evidence of efficient W1282X base correction at the RNA level. A RT-PCR was performed on cells extracted 4 days post-transfection. The amplicons obtained were purified and sent for Sanger sequencing. The chromatograms were analysed through the EditR tool. In this figure are shown the pcDNA, RNP + gRNA A_6 and RNP with gRNA A_7 samples. The gRNA A6 alone, gRNA A7 alone and NG-ABEmax alone analyses displayed results similar to the pcDNA negative control (data not shown). The orange circles show the localisation of the target adenine responsible for the W1282X mutation. The red circle shows the correction from A>G of the W1282X mutation.
This result is the result of a unique experiment (n=1).
Figure 4.5 - Evidence of efficient W1282X base correction at the RNA level

The analysis showed that 50% of the *CFTR* transcripts had the mutation W1282X corrected, in the cells transfected with both ABEmax-NG and gRNA A_7 (Figure 4.5 middle panel). The correction in the DNA was 24%. Regarding the adenine in position c.3847, about 39% of the transcripts contained the modification, compared to 26% for the DNA (Figure 4.4). Those results suggest that there seems to have an accumulation of stabilised RNA for W1282X, but not for R1283G. Those results could be explained by having the W1282X corrected cells stabilised, due to the stop codon not being present anymore. Therefore, with or without the bystander editing, those mRNA not having the stop codon would stabilise. Therefore, if the R1283G editing happened at the same time as the W1282X editing, this could explain the seemingly slightly increased in R1283G editing in the mRNA. However, R1283G alone would not result in a mRNA stabilisation since the stop codon would still be on the mRNA and lead it to degradation through the NMD.

However, the experiment was performed only once, and more experiments would need to be done to confirm those results. If the results are reproducible, a way to discriminate the two assumptions would be to make a cell line containing the R1283G mutation and compare the

RNA stability with a *CFTR* W1282X cell line or a *CFTR* WT cell line, by RT-qPCR. Indeed, if the RNA amount of the R1283G mutant is close to the amount of RNA in the W1282X cells, it would mean that R1283G by itself makes the RNA unstable. However, if the RNA amount is similar to WT cell lines, it would mean that W1282X correction stabilises the RNA even with the R1283G mutation.

4.3 CONCLUSION AND DISCUSSION

The hypothesis tested in this chapter was that the W1282X mutation could be corrected using the ABEmax-NG plasmid with a suitable gRNA. After DNA and RNA analysis, it can be concluded, that base editing successfully corrected the W1282X mutation, with an efficiency of at least 20% using the gRNA A_6. The experiment needs to be reproduced for statistical analyses, but the preliminary results indicated an efficiency about 2-fold higher than what was observed for the AsCas12a HDR editing. However, the efficiency was about the same range of editing than the SpCas9 HDR editing.

Unlike for HDR, no indels were observed for base editing within the limit of detection of the system used (which is estimated to be around ≥1%, depending on how neat the sequencing chromatogram is). Because indels could potentially disrupt the *CFTR* gene permanently, using base editing would substantially reduce those negative side-effects. This makes the base editing technique a potentially safer option for *CFTR* correction. However, as for HDR, off-targets effects will need to be assessed to determine the safety of this technique genome-wide. Even though the DNA was corrected, and it was shown that the correction is also expressed in the mRNA, western blot and functional assays will need to be done to confirm that the correction leads to a functional protein.

A paper from 2020 from the group of Hans Clevers, showed base editing correction of the R785X mutation in *CFTR* using organoids (190). Single cells from organoids were electroporated, and clonal populations were cultured to form organoids that could be tested for CFTR functional correction. Indeed, if the CFTR protein is functionally corrected, forskolin addition would activate CFTR resulting in Cl^- flux into the lumen of the organoids. This Cl^- flux leads to water influx by osmosis, which would make the organoids swell. The paper showed that the organoids derived from the base editing corrected clones responded to forskolin and therefore the base editing corrected CFTR function.

In the case of W1282X, a mixed population of corrected and non-corrected cells are available. Clonal population could be formed, as for Cas12a HDR technique, and cultured on liquid-liquid interface for western blot and Ussing chamber functional analyses. Moreover, western-blot and Ussing chamber of the mixed population could also be interesting, to assess the relevance of 20% *CFTR* correction in a cell population, on the epithelium ion transport. Nevertheless, the results from the group of Hans Clevers are encouraging towards a functional correction of CFTR protein after base editing.

Noticeably, the window of editing could be a limitation for base editing. Indeed, to correct a mutation with base editing, this one needs to be about 14bp upstream the Cas9 PAM sequence. With the ABEmax-NG base editor, the PAM requirement became less stringent compared to the previous editors, making more mutations targetable by base editing. However, every adenine present in the editing window could also be edited, potentially changing the codon sequence and possibly creating a disease-causing mutation.

Therefore, deep characterisation of editing profiles and potential off-target effects will always be needed before any base editing use in animals or humans. Finally, in 2019, the group of Keith Joung showed transcriptome-wide RNA editing from adenine base editor ABEmax (191). Those RNA editing events could be harmful for the organism leading to RNA instability or modified proteins which could lose their function. However, the same group described a new adenine base editor variant called SECURE-ABE, which was shown to reduce the off-target activities on the RNA (see section 1.11.5; (192)).

To optimise genome editing for therapeutic use in Cystic Fibrosis, the correction of a single mutation sounds promising, however, those techniques will be specific for each disease-causing mutation, and might be long and expensive to design and use. However, the development of a technique that would allow the correction of multiple mutations with a single donor could be less expensive and time consuming since only one unique donor and gRNA will be produce for every patient. Moreover, this technique could target many patients, even those with very rare mutations.

To be able to correct multiple mutations in a gene with only one construct, a superexon, or cDNA minigene, can be used. This would have the advantage that many different patients

could be treated with a single set of reagents. The superexon technique was first described in 2006 to correct the *F9* gene in a mouse model of Haemophilia B (170). The construct was a cDNA sequence which contains only the sequence of the exons fused together, giving later its name of superexon (see section 1.12).

In the next part of this thesis, the project focuses on the design and integration into the human genome, of a superexon from exon 23 to 27 of the *CFTR* gene to try and establish proof of principle that this approach can correct multiple CF-causing mutations. This particular superexon is predicted to correct every known *CFTR* mutation from 3849+10kb (intron 22) to 4428insGA (exon 27), comprising mutations W1282X (exon 23) and N1303K (exon 24), which cannot be corrected using the available drugs. The superexon23-27 would target about 4% of CF patients. To avoid deleterious indels induced by Cas9 cutting, the cut site was targeted inside an intron. That way, it was expected that the indels located in the intron would have a less damaging effect. The next two chapters will focus on two ways to integrate the superexon23-27 inside the *CFTR* gene, using NHEJ and HDR repair pathways, with a cut site inside the intron 22 of the gene.

CHAPTER 5 - CORRECTION OF MULTIPLE MUTATIONS WITH A SINGLE DONOR BY HITI

To integrate the superexon in the genome, the **H**omology-**I**ndependent **T**argeted **I**ntegration (HITI) technique was used. The principle of HITI is explained in section 1.10.

5.1 GUIDE DESIGN

The gRNA used for the HITI experiments was a guide designed by David Sanz in Dr. Harrison's laboratory, for a different gene editing strategy that targets the genome in intron

22 of *CFTR*, just upstream the mutation c.3718-2477C>T (3849+10kbC>T). The gRNA was shown to create about 40% indels in transfected cells (193).

5.2 HITI WITH SUPEREXON VERSION 1

5.2.1 - DESIGN OF SUPEREXON VERSION 1 (SEv1)

In order to design a superexon donor, the first objective was to keep the *CFTR* DNA sequences as similar as the genomic sequence as possible to keep the endogenous regulatory sequences. For that reason, the superexon version 1 (SEv1) was designed as followed; the WT *CFTR* cDNA sequence from exon 23 to exon 27 was used. The 102 last nucleotides of the intron 22 were used as splicing acceptor and placed upstream of the cDNA sequence. Those 102 nucleotides were expected to contain the consensus splicing sites (5'- ~18nt Y rich region + NYAG -3', "Y" being T or C nucleotides). Because HITI uses NHEJ pathway, indels could be formed. Therefore, an extra sequence was added upstream to those ~20nt to not disturb the splicing site in the case of deletions happening (yellow box in figure 5.1). Before the splicing site, the gRNA recognition site sequence, in the reverse-complementary orientation compared to the same sequence in the genome, was added (orange arrow in figure 5.1). This gRNA site was flanked with restriction sites to enable future changes if needed. Downstream of the cDNA sequence, the last one hundred nucleotides of the non-coding sequence of exon 27 in the *CFTR* gene were used as a polyadenylation site, containing the consensus sequence 5'- AATAAA + CA + G/T rich region -3'. To prevent polyadenylation site disruption, 35nt were added after the polyadenylation site as a secure sequence for NHEJ.

One nucleotide was modified inside the polyadenylation site, to increase the G/C content (pink box in figure 5.1). Indeed, for the IDT Company to be able to synthesise a DNA sequence, the G/C content needs to be greater than 65%.

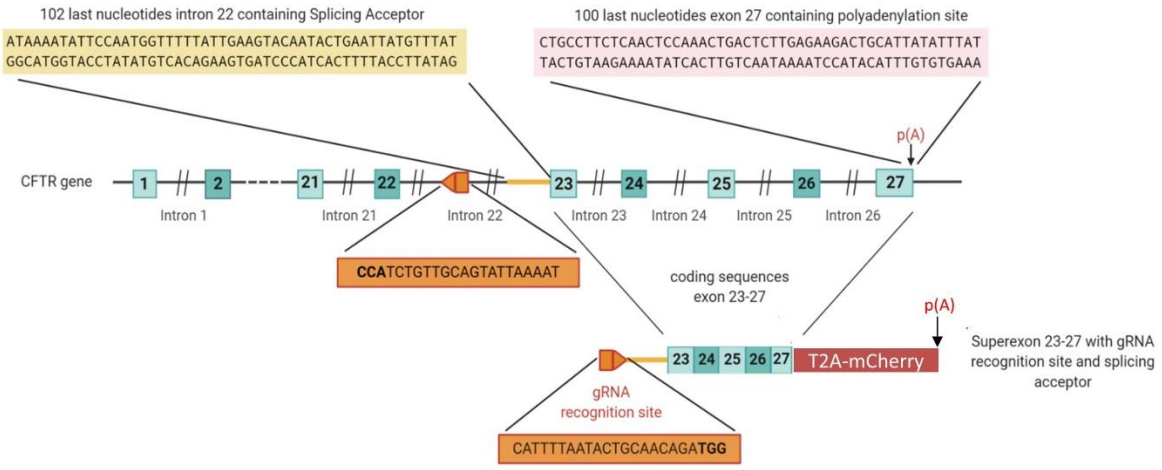


Figure 5.1: Graphical representation of the superexon version 1 design. The wild-type coding sequences of the exons 23 to 27 were fused together to form the superexon23-27. The 102 last nucleotides of the intron 22, containing the splicing acceptor, were added upstream of the superexon (in yellow). Before the splicing site, a gRNA recognition site was added, which was the same sequence as its homologous sequence in the genome, but in reverse-complementary orientation (in orange, with bold sequences being the PAM sites). The 100nt last nucleotides of the exon 27 were used as polyadenylation site (in pink).

Figure 5.1 - Superexon version 1 design

In order to be able to detect easily *CFTR* correction by microscopy, the use of the red fluorescent protein mCherry was used. Indeed, the hypothesis was that the fusion of the mCherry DNA sequence to the *CFTR* gene would lead to the production of red fluorescence only in cells expressing CFTR protein. Therefore, the cells showing red fluorescence in the microscope would be *CFTR*-corrected.

Between the *CFTR* cDNA sequence and the polyadenylation site, the mCherry cDNA sequence, without the first methionine codon, was separated from the *CFTR* sequence via a GT2A sequence. The GT2A cDNA sequence encodes a GSG (Glycine – Serine – Glycine) linker and a T2A oligopeptide sequence (2A peptide sequence derived from the *Thosea asigna* virus). The GT2A sequence allows a co-expression of two distinct proteins from a single transcript.

The T2A sequence is composed of 18 non-conserved amino acids sequence, ending with the consensus sequence **-D(V/I)ExNPG P-** (x being any amino acid, in this case “E”). The division between the two proteins happens between the glycine and the proline of the T2A sequence. Adding a GSG linker before a T2A has been shown to improve the T2A cleavage efficiency (194).

During translation, the ribosome translates the first gene, followed by the GT2A sequence. When the ribosome arrives at the junction glycine/proline, the protein cannot create the peptide bond between the two amino acids. Because of that, the nascent chain is freed, and the ribosome is going to continue its translation from the proline (Figure 5.2).

This GT2A linkage was used to produce a unique mRNA that is going to be processed into two different proteins: CFTR and mCherry, avoiding the mCherry protein from interfering with the CFTR activity.

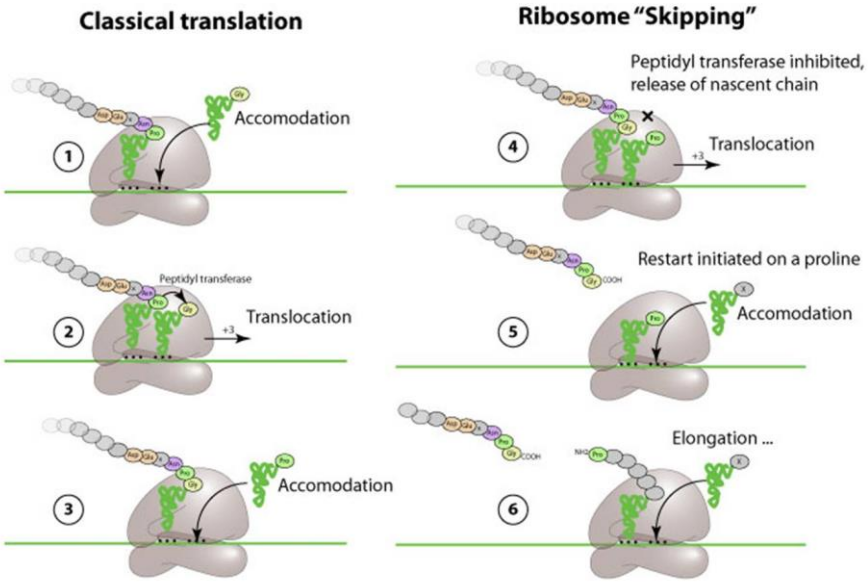


Figure 5.2: Graphic representation of the T2A mechanism. During translation of the mRNA, the prolyl-tRNA is unable to form the peptide bond between the proline and the glycine. Because of this, the peptidyl transferase is inhibited, releasing the nascent peptide. The ribosome will then translocate and restart its translation with the Proline.
***(Image taken from https://viralzone.expasy.org website;** (195))*.

Figure 5.2 - Principle of the T2A mechanism

Finally, two restriction sites in the end of the sequence were added, in order to be able to change the cloning plasmid if necessary. This DNA sequence was ordered as a gBlocks® from IDT.

The plasmid sequence map of the superexon version 1 is shown in figure 5.3.

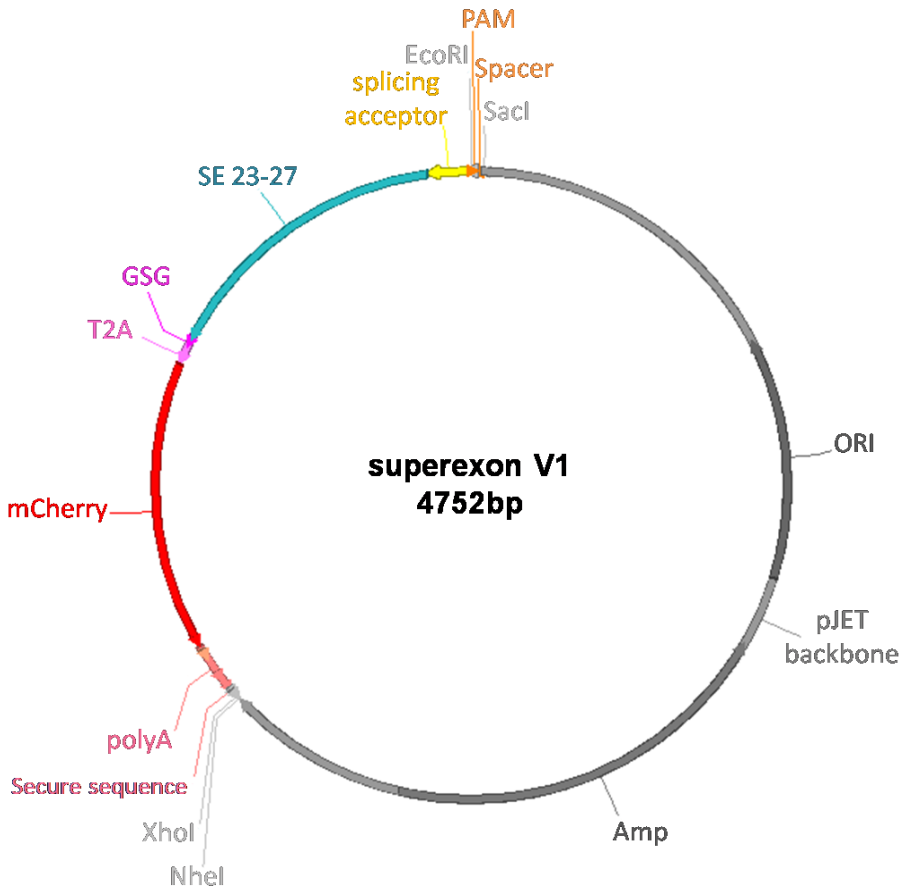


Figure 5.3: The plasmid map of superexon version1.

Figure 5.3 - Plasmid map of the SEv1 construct

The full DNA sequence of the superexon version 1 without the backbone is available in appendix, figure S1.

5.2.2 - ANALYSIS OF THE TRANSFECTION EFFICIENCY

The first goal of this chapter was to assess if it was possible to integrate a superexon using the HITI system into the *CFTR* gene. HEK293 cells were used because those cells are easy to transfect using plasmids. However, because HEK293 cells do not express CFTR (173), this preliminary study was limited to assessing superexon integration in the gene, and optimising transfection.

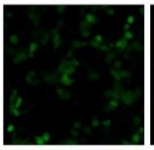
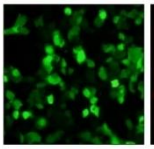
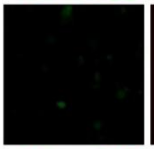
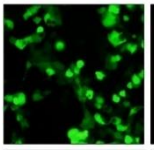
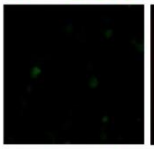
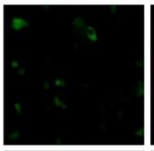
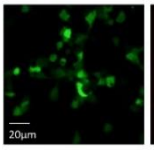
HEK293 cells were transfected using Lipofectamine 2000 with either pcDNA empty plasmid, SEv1 donor alone and Cas9/gRNA plasmid alone, as controls, or with both Cas9/gRNA and SEv1 plasmids.

The gRNA and SpCas9 were encoded by the pSpCas9(BB)-2A-GFP plasmid expressing GFP. Different ratios of Cas9/gRNA:SEv1 (1:1; 1:3 and 3:1) and different amount of plasmids (500ng; 750ng; 1500ng) were used (Figure 5.4A).

Forty-eight hours post-transfection, the cells were observed under an Olympus IX70 fluorescent microscope, with a 40x magnification, for expression of GFP (from gRNA/Cas9 plasmid) and mCherry (integration and expression of CFTR-GT2A-mCherry protein from SEv1). The software used to capture images was the Cell A Imaging software (Olympus).

The microscopy results with each condition are shown in figure 5.4A. For each condition, between 4 and 12 images were taken in order to estimate the efficiency of correction (4 images for pcDNA control, 6 images for SEv1 alone, 7 images for Cas9/gRNA alone, 9 images for the both 1:1 ratios, 8 images for the both SEv1:Cas9/gRNA 1:3 ratios as well as the 3(1.375ng):1(0.125ng) ratio, and 12 images were analysed for the SEv1:Cas9/gRNA 3(1125ng):1(0.375ng)). For each image, the number of green cells and red cells was reported, and the efficiency of correction corresponded to the percentage of green cells that were also red (Figure 5.4B).

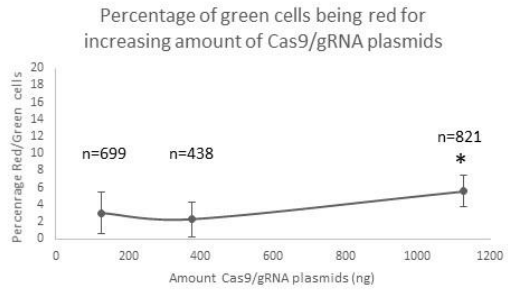
A)

# images	GFP	mCherry	Merge	Ratio SEv1:gRNA	SEv1	Cas9/gRNA	pcDNA
n=4				N/A	0	0	1500ng
n=6				N/A	750ng	0	750ng
n=7				N/A	0	750ng	750ng
n=9				1:1	750ng	750ng	0
n=9				1:1	375ng	375ng	750ng
n=8				1:3	375ng	1125ng	0
n=8				1:3	125ng	375ng	1000ng
n=8				3:1	375ng	125ng	1000ng
n=12				3:1	1125ng	375ng	0

B)

SEv1 = 375ng			
amount gRNA	125ng	375ng	1125ng
efficiency	3.05	2.29	5.59
SD	2.46	2.05	1.86

gRNA = 375ng			
amount SEv1	125ng	375ng	1125ng
efficiency	0.70	2.29	13.04
SD	1.14	2.05	3.24



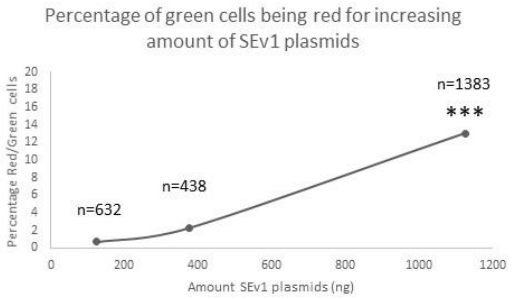


Figure 5.4: Fluorescent microscopy results and analyses for different ratios and amounts of gRNA/Cas9 and SEv1 plasmids. HEK293 cells were transfected using Lipofectamine 2000, with either pcDNA empty plasmid alone as a negative control, SEv1 plasmid alone, Cas9/gRNA plasmid alone, or different ratios and amounts SEv1:Cas9/gRNA. Forty-eight hours post-transfection, each conditions were observed in the Olympus fluorescent microscope at a 40x magnification. A) Example of microscopy images for each conditions described in the table. The gRNA/Cas9 plasmid expressed GFP while the SEv1 plasmid expressed mCherry only if integrated under the control of a promoter. Scale bar = 20µm. The number of images analysed for each conditions in shown in the left column.
B) Graphical representations of the correction efficiency analyses performed on the microscopy images. The graph on the left shows the percentage of red cells for increasing amounts of Cas9/gRNA plasmids, and the graph on the right shows the percentage of red cells for increasing amounts of SEv1 plasmids transfected.
The n numbers correspond to the number of green cells analysed for each condition. The Ttest statistics results showed significant data only for the last sample of each graph (p=0.03(*) and p=2.07E-08(***)).

Figure 5.4 - Fluorescent microscopy analyses for different ratios and amounts of gRNA/Cas9 and SEv1 plasmids

As expected, figure 5.4A displayed presence of GFP only in cells transfected with the Cas9/gRNA plasmid. Moreover, the percentage of cells expressing GFP appeared to increase with the increasing amount of plasmid transfected (Figure 5.4A and B).

However, it was surprising to see mCherry expression, since first, the superexon-mCherry plasmid does not contain any promoter to induce mCherry production, and second, if the superexon-mCherry plasmid integrated into the intron 22 of the *CFTR* gene, as predicted, no expression would be expected as the endogenous *CFTR* promoter is not active in HEK293 cells. Noticeably, many but not all of the red cells were also green, suggesting that in order to have mCherry production, the expression of the pSpCas9(BB)-2A-GFP plasmid was not always required.

The analyses done on the microscopy images showed that the amount of SEv1 plasmid transfected in the cells seemed proportional to the percentage of red cells and therefore to the percentage of *CFTR* correction. However, the amount of pSpCas9(BB)-2A-GFP plasmid did not seem to have a significant impact on the correction efficiency. Importantly, it can be noticed that the best correction efficiencies found in this experiment were between 10 and 16% which are high compared to what was shown on CF previously for superexon.

It is however important to note that the efficiencies were determined by counting the cells on the microscopy images, based on assigning if the cells were green/red or not, which can be somewhat subjective. Therefore, a confirmation with a fluorescence-activated cell sorting (FACS) analysis, where the fluorescence intensity is measured directly by the cytometer without human bias, would be needed in order to further characterise mCherry expression.

5.2.3 - SUPEREXON INTEGRATION INTO THE *CFTR* GENE

Before doing the FACS analyses, it was decided to confirm first, that the integration of the superexon into the *CFTR* gene had occurred. HEK293 cells were transfected, using Lipofectamine 2000, with either pcDNA empty plasmid alone as a negative control, the SEv1 plasmid alone, the Cas9/gRNA plasmid alone or both SEv1 + Cas9/gRNA plasmids together at a 1:1 ratio. Forty-eight hours post-transfection, the DNA was extracted, and a PCR was done using a set of primers targeting the intron 22 before the gRNA recognition site (3849+10kb FW) and exon 23 (Superexon 5' REV). With those primers, if the superexon has been integrated in the intron 22 of the *CFTR* gene, the expected band should have a size of 446bp. However, if the superexon has not been integrated, the expected band would have a size of 2827bp (Figure 5.5A). The PCR results are shown in figure 5.5B.

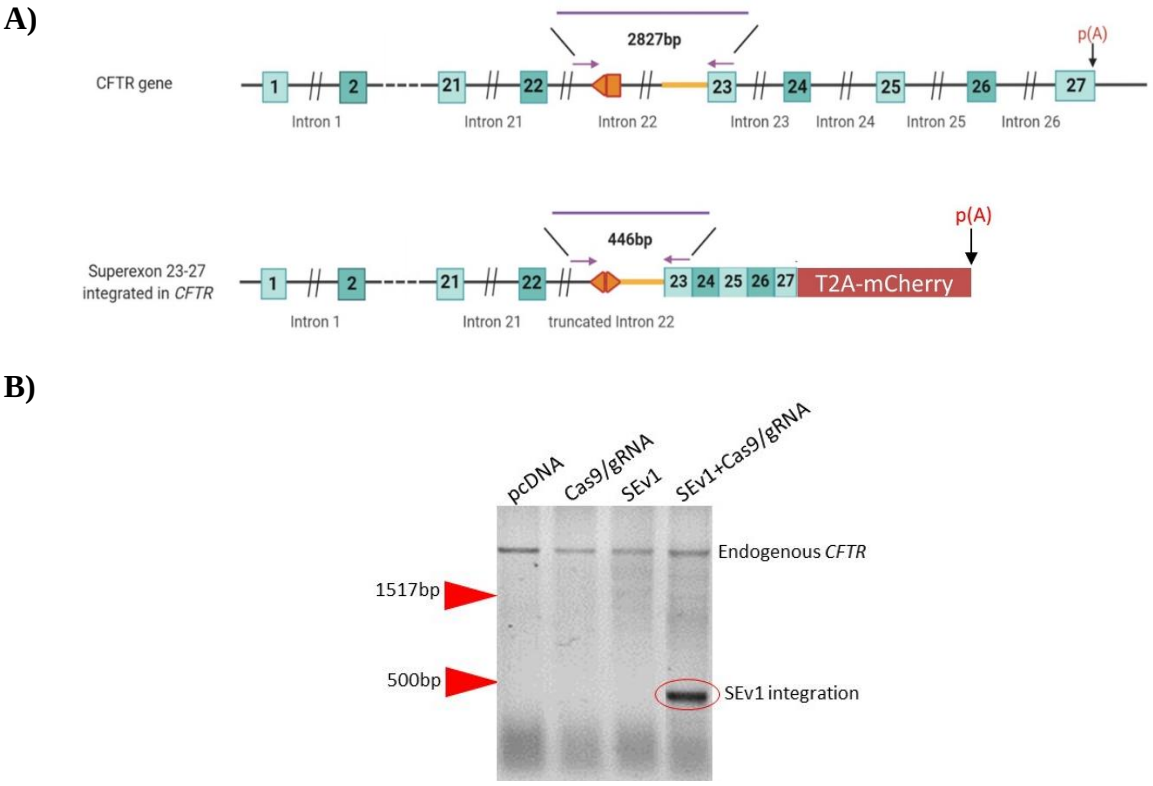


Figure 5.5: Evidence of superexon integration in HEK293 cells. HEK293 cells were transfected, using lipofectamine 2000, with either pcDNA empty plasmid alone as a negative control, Cas9/gRNA plasmid alone, SEv1 plasmid alone or SEv1 + Cas9/gRNA at a 1:1 ratio. Forty-eight hours post-transfection, the DNA was extracted, and a PCR was performed around the 5' junction. A) Graphic representation of the primers' position and the expected sizes for both endogenous CFTR and superexon-integrated PCR products. B) PCR results showing a band of expected size for superexon integration, only when both the SEv1 and Cas9/gRNA plasmids were co-transfected.

Figure 5.5 - Evidence of superexon version 1 integration in HEK293 cells

The PCR results show presence of the endogenous genomic *CFTR* in both samples. Because the sample transfected with gRNA and SEv1 plasmids is a pool of cells containing some transfected and non-transfected cells as well as probably heterozygous corrected cells, the presence of WT *CFTR* gene was expected.

Interestingly, only the sample transfected with both the gRNA and SEv1 plasmids contains the smaller band corresponding to integration of the superexon into the intron 22 of the *CFTR* gene. This result is a good indication of the superexon integration into the *CFTR* gene.

Now that the superexon integration has been confirmed, the next step was to investigate the mCherry data obtained from microscopy by FACS analysis.

5.2.4 - ANALYSIS OF THE mCHERRY-POSITIVE CELLS

5.2.4a - FACS analysis and sorting

To perform the FACS analysis, new HEK293 cells were transfected with pcDNA empty plasmid, Cas9/gRNA plasmid only (750ng), SEv1 plasmid only (1125ng), or both Cas9/gRNA and SEv1 plasmids at ratios 1:1 (750ng:750ng) or 1:3 (375ng:1125ng). Forty-eight hours post-transfection, the samples were run through a BD FACS Aria™ Fusion cell sorter, for GFP and mCherry analysis. The mCherry-positive cells were sorted and kept in culture to form an mCherry-positive enriched cell population. The results of the FACS analysis are shown in figure 5.6.

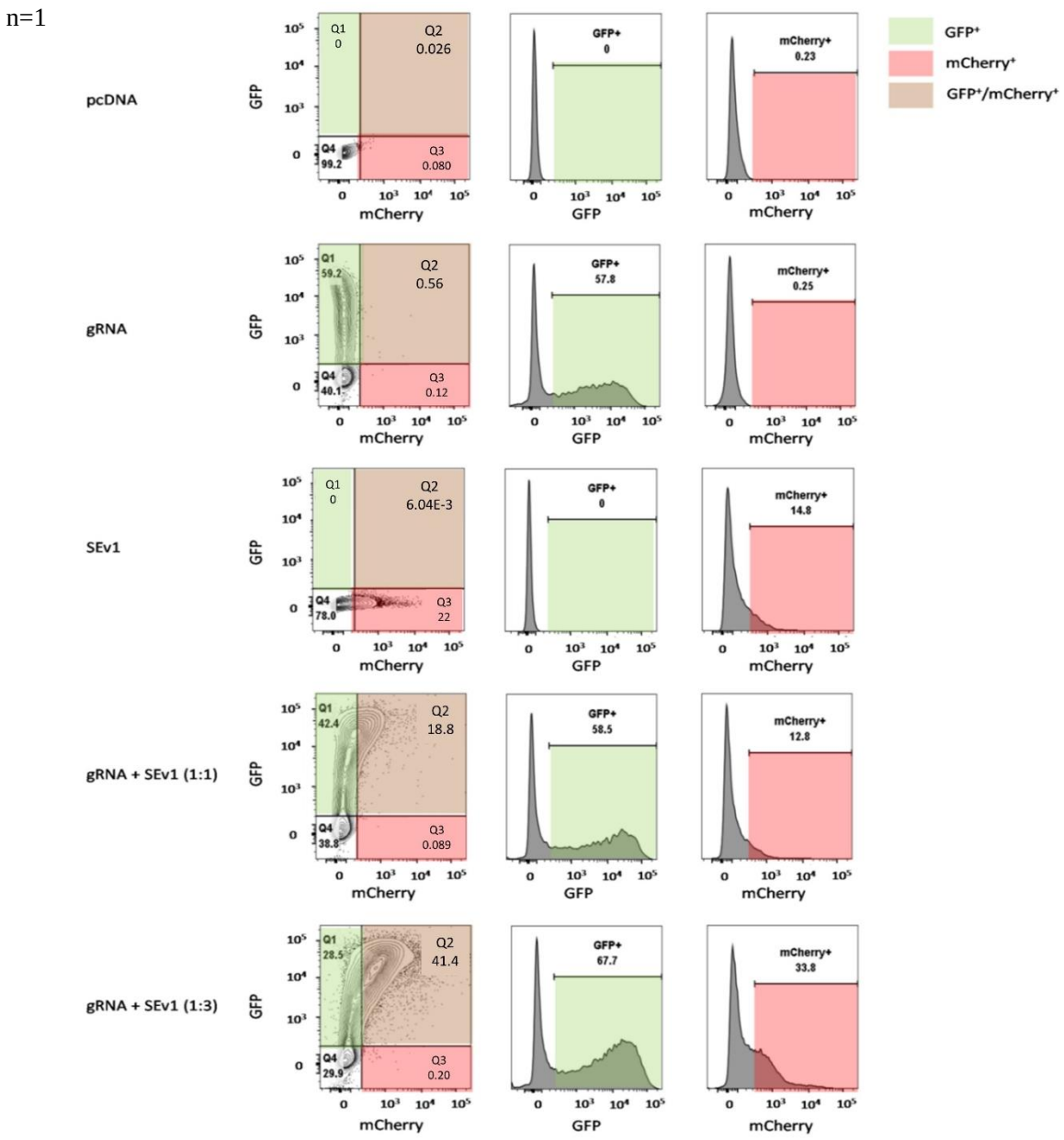


Figure 5.6: FACS analyses showing mCherry expression in cells transfected with the SEv1 plasmid. HEK293 cells were transfected using Lipofectamine 2000, with either pcDNA empty plasmid as a negative control, Cas9/gRNA only, SEv1 donor only and both Cas9/gRNA + SEv1 at ratios 1:1 and 1:3. Forty-eight hours post-transfection, the cells were analysed through a cell sorter for GFP and mCherry expression. GFP expression comes from the Cas9/gRNA plasmid which contains a GFP cassette, while mCherry comes from integration and expression of the SEv1-GT2A-mCherry construct into the CFTR gene. The dot plots represent cells that are GFP-positive (in the green quarter), mCherry-positive (in the red quarter), negative for both proteins (in the white quarter) or positive for both proteins (in the brown quarter). Each quarter contains the percentage of cells belonging to its category. Histograms on the right display the percentage of GFP-positive cells (on the left) or mCherry-positive cells (on the right), amongst the cell population. The results obtained were from a unique experiment (n=1).

Figure 5.6 - FACS analyses showing mCherry expression in cells transfected with the SEv1 plasmid

Dot plots in figure 5.6 show percentages of GFP and mCherry-positive cells. The number of green cells corresponds to the number of cells, which were successfully transfected with the pSpCas9(BB)-2A-GFP plasmid and expressed GFP. Looking at the condition Cas9/gRNA alone, the percentage of green cells would give an approximation of the total transfection efficiency. From the gRNA alone dot plot in figure 5.6, it can be estimated that the transfection efficiency was about 60%.

Second thing to note is the 22% of mCherry-positive cells when transfecting the cells with 1125ng of SEv1 plasmid only. Assuming a similar transfection efficiency of about 60%, this result initially suggested that about 1/3 of those cells were expressing mCherry. Since the SEv1 construct contains no known promoter, it was initially difficult to explain this high level of mCherry expression.

In theory, the expression could be explained if the SEv1 plasmid had undergone an integration event that enabled expression of the CFTR-mCherry sequences encoded by the superexon. However, because there was no Cas9/gRNA, neither the genome nor the donor could be specifically cut, suggesting that the mCherry expression was not due to HITI integration. It is possible that the superexon could be integrating either spontaneously or via homology-directed events. Indeed, the superexon and the genome shared some homologous sequences such as the exon sequences or the 102bp ending the intron 22 and containing the splicing acceptor. However, it was surprising that the background level of mCherry was so high, and this is discussed in more details in the discussion for this chapter.

In cells where both the Cas9/gRNA and SEv1 plasmids were co-transfected, an increase in mCherry-positive cells can be seen with the increased amount of superexon transfected. This data concurs with the results described previously by microscopy. Moreover, it can be noted that with less Cas/gRNA plasmid (325ng of plasmid), the percentage of GFP positive cells could reach up to 70%, showing that the amount of pSpCas9(BB)-2A-GFP plasmid did not seem to affect the percentage of green cells significantly.

Comparing the conditions SEv1 alone with SEv1:Cas9/gRNA (1125ng:375ng), a 2-fold increase in the percentage of mCherry-positive cells can be noticed (22% and 42% respectively). Moreover, it can also be noticed that the mCherry seemed to be brighter in the latter condition (Figure 5.6, dot plots and mCherry histograms).

Following the FACS results, it appeared that when the cells were transfected at a Cas9/gRNA:SEv1 1:3 ratio, the percentage of CFTR correction, showed by mCherry expression, could reach up to 40% which suggests a very good HITI efficiency. Importantly, if only the transfected cells were taken into account, instead of the whole cell population, the correction efficiency would be about 70%.

However, as stated in section 5.4.1, HEK293 do not have a *CFTR* active promoter, therefore, those results are unexpected. In the conclusion and discussion part of this chapter, some assumptions explaining mCherry expression will be discussed.

To further investigate the mCherry expression, it was decided to isolate and culture the mCherry-positive cells from the condition Cas9/gRNA + SEv1 (1:3) to create an enriched mCherry-positive cell population.

During culture, cells were observed in the fluorescent microscope at different time points, to monitor mCherry enrichment (Figure 5.7).

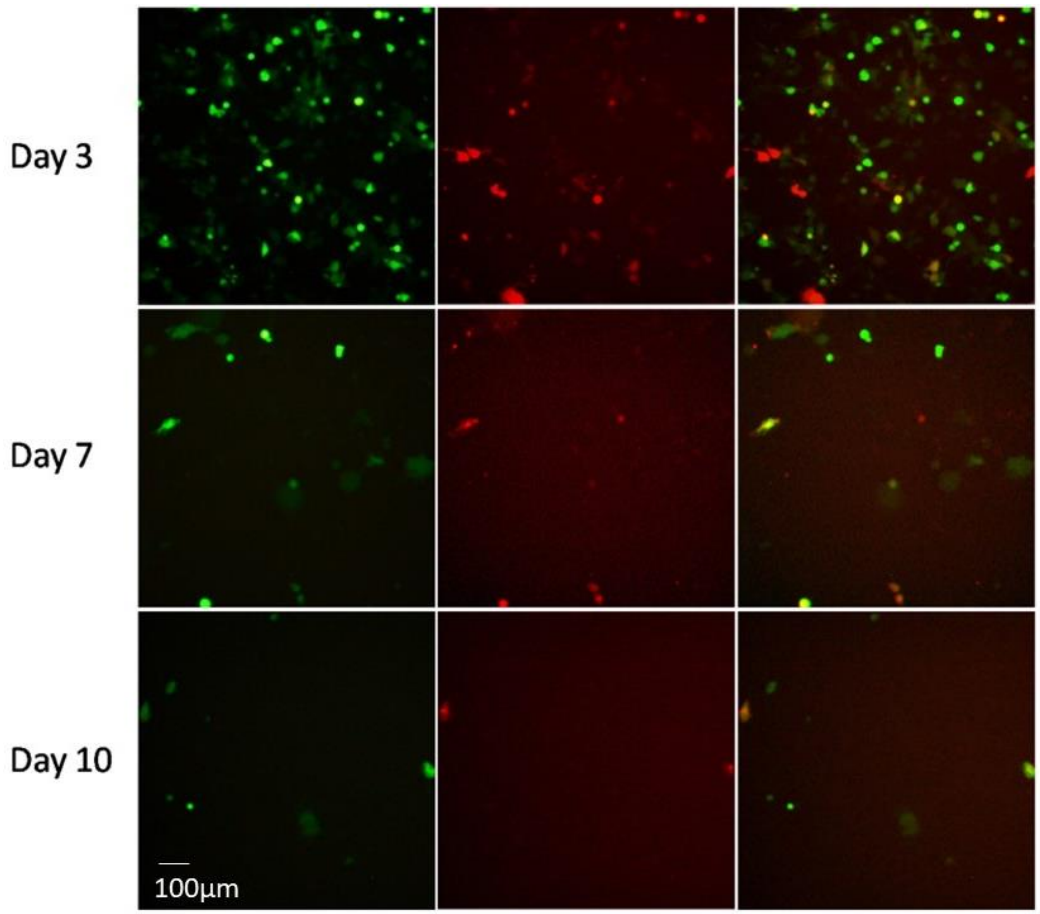


Figure 5.7: Microscopy of mCherry-enriched cells (magnification 20x) at day 3, day 7 and day 10 post-cell sorting. Sorted cells were observed on the Olympus fluorescent microscope. Both mCherry and GFP disappeared with time. Scale bar = 100µm.

Figure 5.7 - Microscopy of mCherry-enriched cells at day 3, day 7 and day 10 post-cell sorting

As shown in figure 5.7, both GFP and mCherry expression disappeared with time. Interestingly, it looks like both signals disappeared at the same time. Whilst this is to be expected for the GFP signal, due to either silencing or loss by dilution of the pSpCas9(BB)-2A-GFP plasmid, the loss of mCherry signal is difficult to reconcile with permanent integration into the genome by the HITI process.

It was then decided to confirm integration of SEv1 into the *CFTR* gene, in those sorted cells, by sequencing.

5.2.4b - SEv1 integration into the *CFTR* gene, through HITI, in HEK293 mCherry-enriched cells

Once the sorted cells reached confluency (14 days post-cell sorting), the DNA was extracted, and a PCR was done using the same set of primers as in section 5.2.3. The negative pcDNA sample and the positive gRNA + SEv1 sample from section 5.2.3, were used as control (Figure 5.8A).

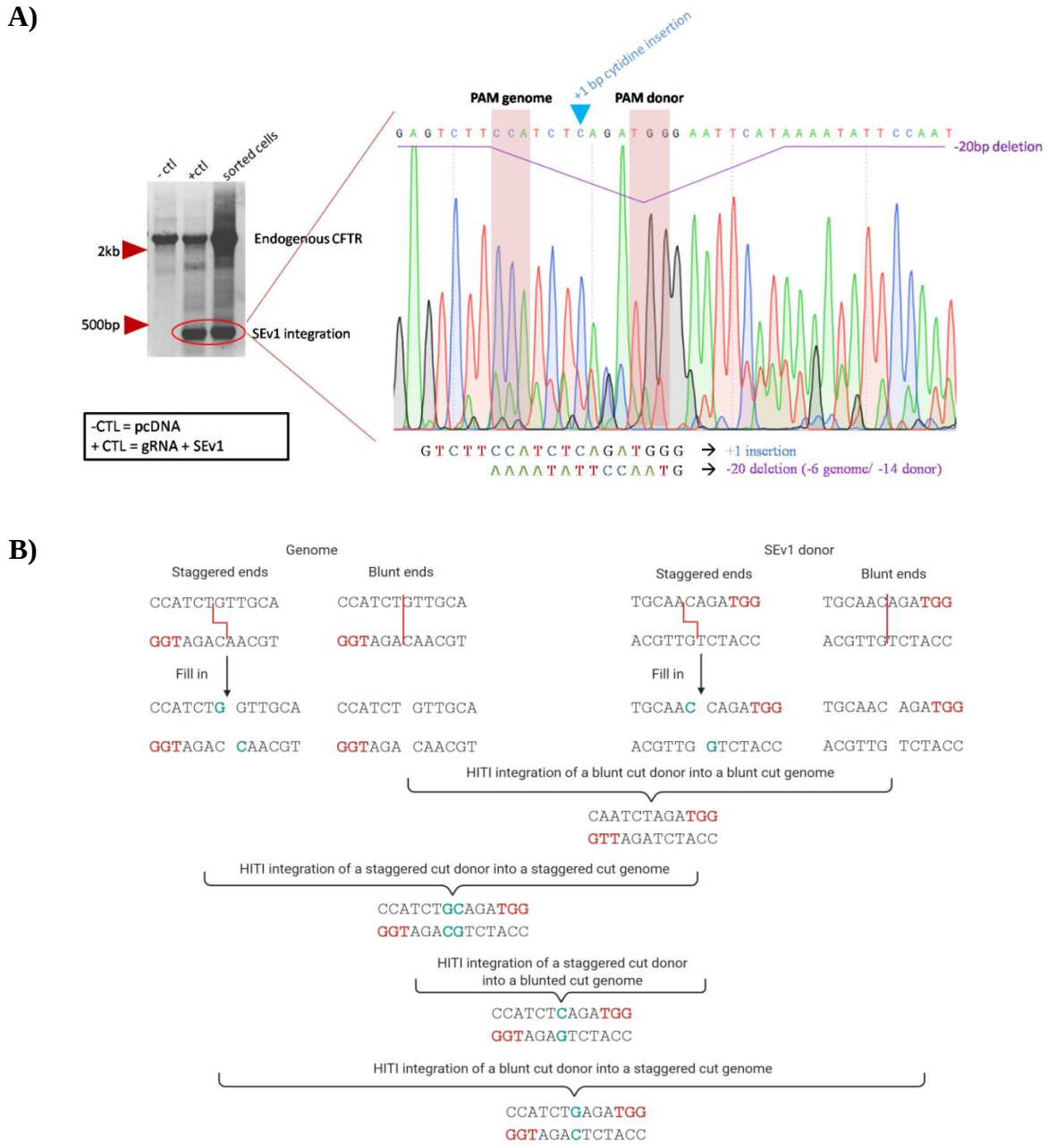


Figure 5.8: Evidence of superexon integration through HITI, in the mCherry-positive enriched HEK293 cells. Fourteen days post-cell sorting, the DNA of the cells was extracted, and a PCR was done around the 5' junction, using a previous pcDNA negative control and SEv1 + Cas9/gRNA positive control. (A) PCR results of the 5' junction integration. A band of expected size for SEv1 integration can be observed only in samples transfected with both the Cas9/gRNA and the SEv1 plasmids. The small band was sequenced and shown on the left. The sequencing results demonstrates the junction between the genomic CFTR and the donor SEv1. The HITI integration is showed by the presence of the two PAM sequences joined together. Two different sequences can be observed in the chromatogram and are reported below it. The major sequence contained a +1 cytidine insertion, while the other contained a -20nt deletion. B) Prediction of nucleotide insertion if SpCas9 created one nucleotide 5' staggered ends in either the genomic DNA, the donor plasmid or both.

Figure 5.8 - Evidence of superexon integration through HITI, in the mCherry-positive enriched HEK293 cells

The PCR results showed a 446bp band in the HEK293 mCherry-enriched cells, corresponding to the integration of the SEv1 into the intron 22 of *CFTR*. The band was sent for Sanger sequencing to confirm the junction between the *CFTR* genomic sequence and the SEv1 donor (Figure 5.8A). The sequencing confirmed integration of the donor plasmid at the expected localisation in intron 22 of *CFTR* by HITI. The HITI integration could be confirmed with the presence of the two PAM sequences from both the genome and the donor plasmid, joined together.

Interestingly, a (+1bp) insertion of a cytidine and a less abundant (-20bp) deletion can be observed in the junction between the genome and the donor (Figure 5.8A). A paper from 2018 showed evidence of a (+1bp) insertion being the most common indel after Cas9 cutting. Interestingly, they also showed that this insertion could be predicted, since it would come from the SpCas9 doing a staggered cut 4nt upstream the NGG PAM on the non-target strand, while cutting 3nt upstream the PAM on the target strand (96). Following that information, it is possible to predict if the cytidine observed was the insertion expected if SpCas9 was creating a one nucleotide staggered break (Figure 5.8B). In figure 5.8B is shown every possibility of insertion if SpCas9 cut the genome or/and the SEv1 donor in a blunt or staggered manner. Interestingly, there could be four possibilities of indels depending whether the DNA underwent either a blunt or a staggered cut. However, the insertion of a cytidine would only be present if the SEv1 donor underwent staggered cutting while the genome underwent blunt cutting. Therefore, if the staggered cut happened mostly in the donor and not as much in the genome, it could be predicted that the most common indel would be a (+1bp) cytidine. However, why the donor would undergo more staggered cut than the genome is unknown.

Another possibility could be that the mCherry enrichment was composed of very few HITI integrated SEv1 donor, explaining the red disappearing. It is possible that this (+1bp) cytidine insertion and the (-20bp) deletion were the result of unique integrated cells that multiplied in culture. Therefore, it could be assumed that if more integrations happened, maybe the four different kind of indels predicted would be seen in the Sanger sequencing profile.

Finally, to confirm if the mCherry was indeed coming from a possible CFTR expression, RT-PCRs were done by a colleague in the lab. However, no mRNAs were detected in those mCherry-enriched HEK293 cells (data not shown), suggesting that the mCherry was not coming from CFTR expression.

5.2.5 - CONCLUSION AND DISCUSSION

As stated earlier, the major goal of this chapter was to determine if HITI integration was possible and identify the best conditions to achieve this. As detailed in this chapter, correct integration of SEv1 by HITI was confirmed by PCR and sequencing, and a 3:1 ratio of SEv1:Cas9/gRNA was shown to be most suitable. No expression of mCherry was expected in successfully edited cells as CFTR is not normally expressed in HEK293 cells. However, as described above, considerable mCherry was observed but it was not stable over time, even though HITI integration was still detected after fainting of the mCherry signal.

Here, several theories to possibly explain the mCherry expression are discussed:

First of all, it is known that HEK293 cells do not express CFTR. One theory was that, somehow, the integration of the superexon activated the *CFTR* gene transcription, maybe through opening of the chromatin, or by removing some negative regulators that could be localised in the missing introns (in23 to in26) or in the missing 3'UTR. However, this expression would be silenced overtime. The silencing could be due to DNA methylation, histone modifications, chromatin condensation or intrinsic regulation (e.g. through miRNA).

The expression of mCherry in HEK293 cells was quite puzzling until the paper from d'Adda di Fagagna's group was analysed (196). The paper described the formation of long non-coding RNA formed when a plasmid is cut or broken inside the cells (196). The group designed a plasmid containing an I-SceI restriction site, without promoter or poly-adenylation sites. They described no RNA production from this plasmid. However, when they added the I-SceI enzyme to create a DSB, they detected a divergent bidirectional RNA synthesis originating from the DSB (Figure 5.9; (196)).

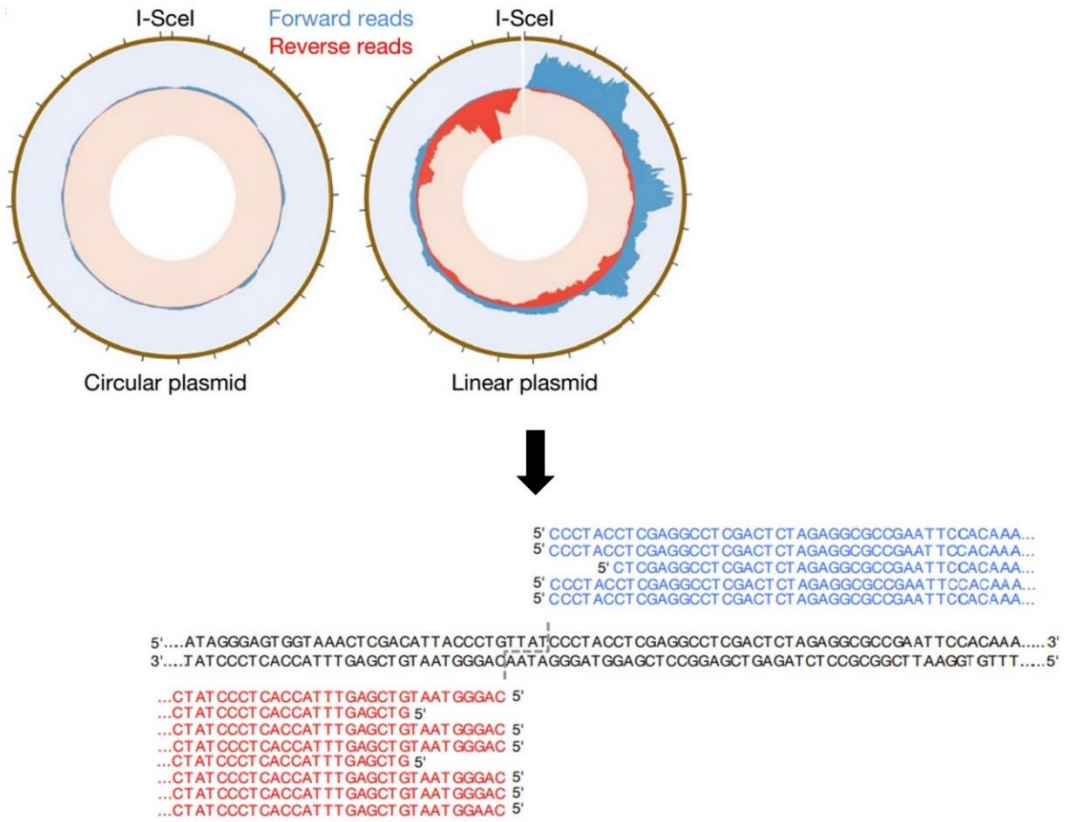


Figure 5.9: Double strand break in a plasmid leads to divergent bidirectional RNA synthesis. When the plasmid does not have a DSB, only a background of RNA synthesis could be observed. When the DSB was created using I-SceI, a large amount of RNA synthesis could be observed originated from the DSB and towards both direction (5'→3' in blue and 3'--\|5' in red). ***The figure was adapted from Michelini et al. 2017*** (196).

Figure 5.9 - Double strand break in a plasmid leads to divergent bidirectional RNA synthesis

Indeed, in response to plasmid double-strand break, damage-induced long non-coding RNAs (lncRNAs) are produced by the RNA polymerase II as part of the DNA damage response (196). Following this paper, a possible explanation is that the mCherry signal was actually coming from the SEv1 plasmid itself. This could explain the loss of mCherry by dilution or silencing at the same time as the GFP plasmid. It could also explain the absence of *CFTR* mRNA and the proportionality between the amount of plasmid transfected in the cells and the amount of mCherry-positive cells observed in both microscopy and FACS results. This phenomenon could explain why, in the FACS data, there were some mCherry (~22%) produced when the plasmid was delivered by itself (resulting from spontaneous DNA break), and a higher mCherry-positive cell number when the plasmid was co-transfected with the Cas9/gRNA plasmid (42% for ratio gRNA: SEv1 (1:3)).

However, the pJET plasmid backbone does not have any known promoter sequences that could express mCherry, so how is it being expressed?

The paper from 2017, the RNAs produced were non-coding (196). Importantly, the plasmid that they used did not have any promoter or polyadenylation site, therefore no proteins could be produced from those lncRNA. However, the SEv1 plasmid construct containing mCherry had a coding sequence with a polyadenylation site, allowing the potential formation of a stable mRNA. Even though there was no 5'-ATG-3' start codon inserted in the mCherry construct, there are several other 5'-ATG-3' available in the sequence, which the cell could use as a start codon if a suitable Kozak sequence is present. Therefore, it could be possible that the RNA produced through SEv1 could encode a translatable RNA.

A way to test this hypothesis would be to create two mCherry plasmids with and without a polyadenylation site, without a superexon and with a plasmid specific gRNA/Cas9 recognition site. If the plasmid without polyadenylation site does not show any mCherry with or without cut, but the plasmid with a polyadenylation site shows mCherry expression with more expression after cut, it could confirm the model proposed.

Even though the mCherry expression cannot be definitely explained, the integration of the superexon, however, was confirmed by PCR and sequencing. The next step was to see if there would be RNA expression. However, HEK cells do not express CFTR, therefore, another cell type was needed to further study the HITI integration. To do so, 16HBE14o$^-$ cells were used.

This cell line was used because of the availability of a W1282X mutant cell line, in which the mutation can be corrected with the superexon23-27. However, because those cells already expressed the endogenous CFTR, a new superexon design was necessary to differentiate between transcripts from the endogenous CFTR and transcripts coming from the superexon integration, by RT-PCR. Therefore, the superexon version 2 was designed and used for further experiments.

5.3 HITI WITH SUPEREXON VERSION 2

5.3.1 - DESIGN OF SUPEREXON VERSION 2 (SEv2)

The superexon version 2 (SEv2) was designed as followed: First, the mCherry sequence was removed to shorten the overall donor sequence. A smaller construct is expected to slightly increase the integration efficiency (197, 198). To differentiate superexon integration and expression over the endogenous CFTR in cells, and to prevent eventual homology-dependent integration events, a codon optimised sequence of exon 23 was designed using the JAVA Codon Adaptation Tool (199). This software designs the codon-optimised sequence for the human genome of any inputted sequences.

A modified intron 22 sequence was added before the splicing acceptor and flanking the gRNA sequence. In addition, a PCR reverse primer sequence was added between the modified intron 22 sequence and the splicing acceptor (Figure 5.10A). Those two factors were designed to obtain a PCR product of a similar size between the endogenous *CFTR* gene and the SEv2-integrated construct, for deep sequencing, using the same sets of primers. Indeed, for deep sequencing experiments, amplicons' sizes for non-edited *CFTR* and edited *CFTR* need to be almost identical, to avoid amplification bias. With the SEv2 construct, and with a forward primer targeting the intron 22 before the genomic cut, the outcome of HITI integration should result in an amplicon size of 145bp (Figure 5.10B) while the same primers set used in WT sequence would give an amplicon of 139bp (Figure 5.10A).

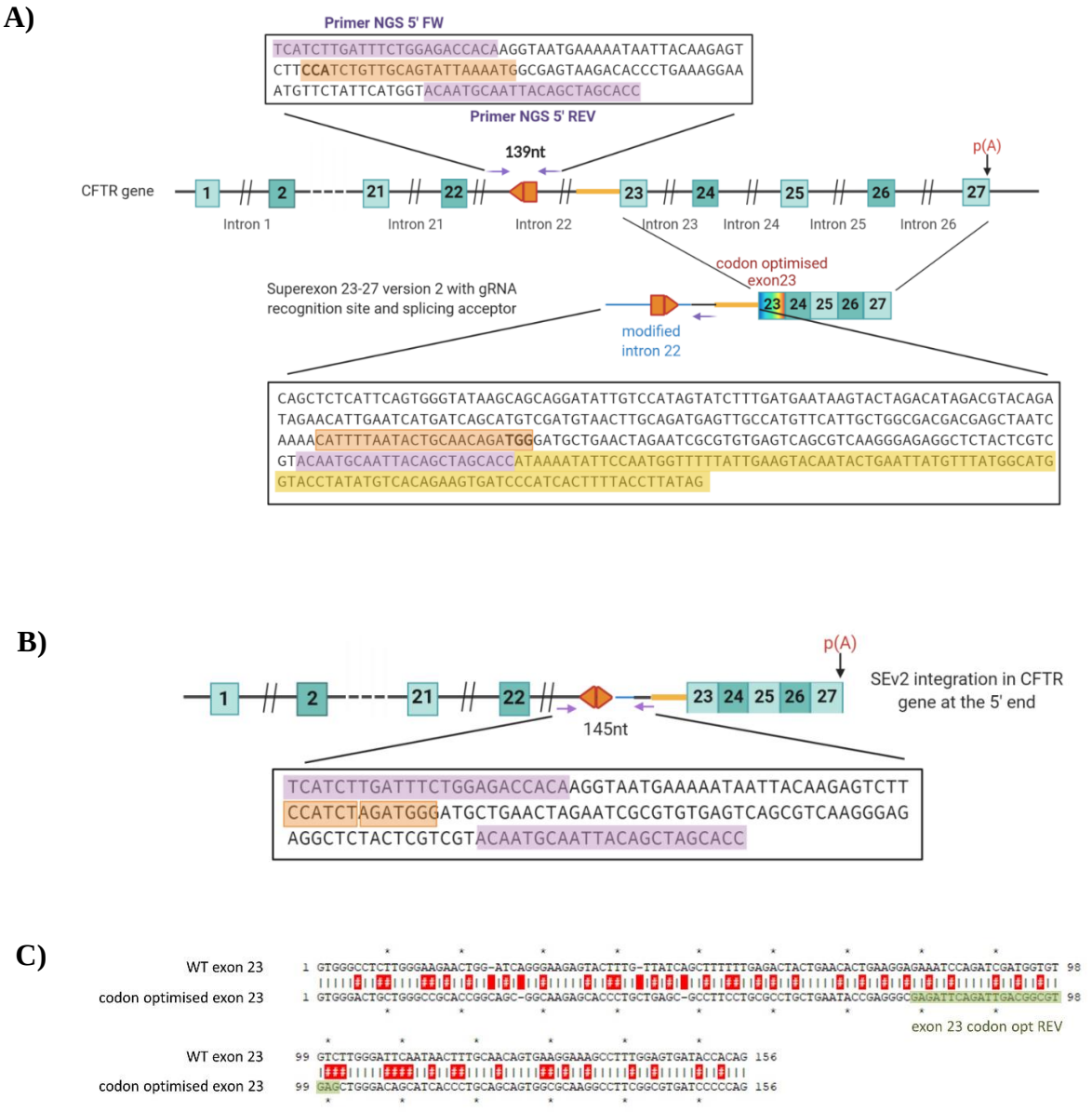


Figure 5.10: A) Graphic representation of SEv2, showing the CFTR cDNA sequence from exon 23 to exon 27, with a codon optimised exon 23 to differentiate from the endogenous CFTR before and after expression. Around the gRNA recognition site, a short, modified sequence of the intron 22 was added for NGS analysis (blue line). The DNA sequence on the top, represents the expected PCR amplicon sequence obtained after deep sequencing. Between the gRNA recognition site and the splicing acceptor, a 23nt reverse primer identical to a sequence in WT intron 22 was added (purple arrow and purple sequence). B) Graphic representation of the SEv2 integration into the CFTR gene at the 5' end, with the expected NGS PCR product sequence. C) Sequence alignment between the WT exon 23 and the codon optimised exon 23.

Figure 5.10 - Superexon version 2 design

The SEv2 was ordered as a plasmid from IDT, in the pUCIDT(Amp) backbone. The SEv2 plasmid map is shown in figure 5.11.

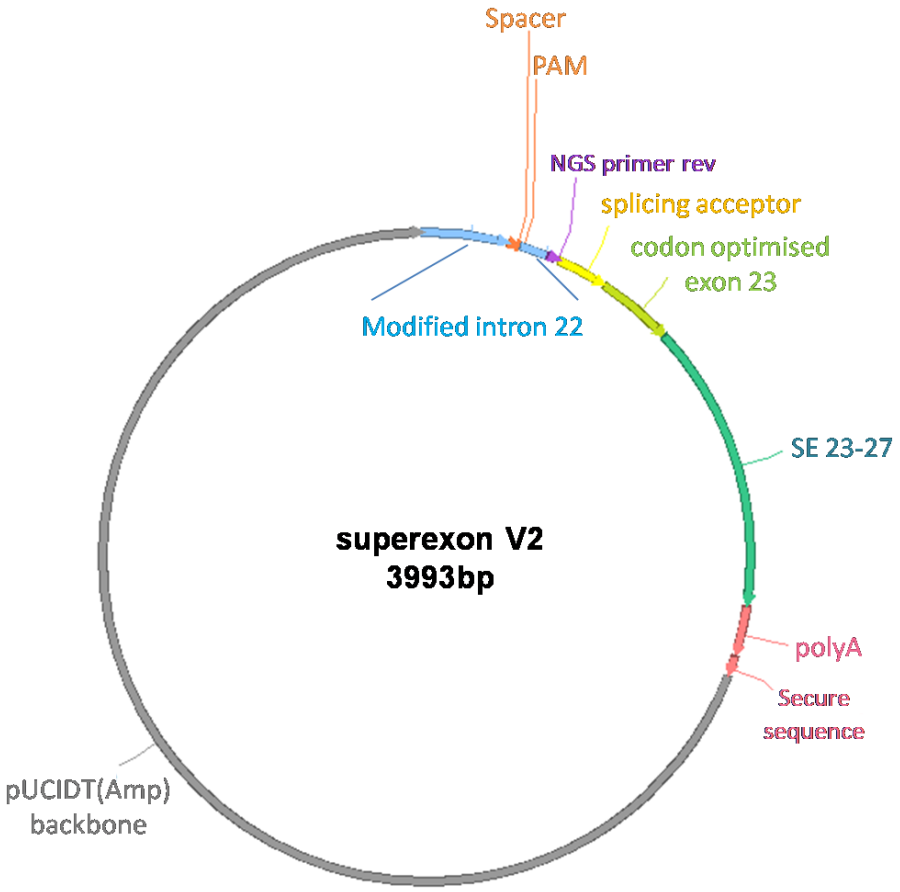


Figure 5.11: The plasmid map of the superexon version 2 donor plasmid design.
Figure 5.11 - Plasmid map of the SEv2 construct

The full DNA sequence of the superexon version 2 without the backbone is available in appendix, figure S2.

5.3.2 - INTEGRATION OF SEV2 INTO THE *CFTR* GENE OF HEK293 CELLS

Before analysing the superexon integration efficiency and expression, it was important to first confirm that the new superexon construct could also integrate.

HEK293 cells were transfected with either pcDNA empty plasmid alone as a negative control, the plasmid SEv2 alone, the plasmid pSpCas9(BB)-2A-GFP (Cas9/gRNA) alone, or both plasmids together. After 24 hours incubation, the cells were observed in a Leica DMI 3000 B fluorescent microscope, using the Leica Application Suite software (LAS v4.1) to capture images, in order to estimate the transfection efficiency (Figure 5.12).

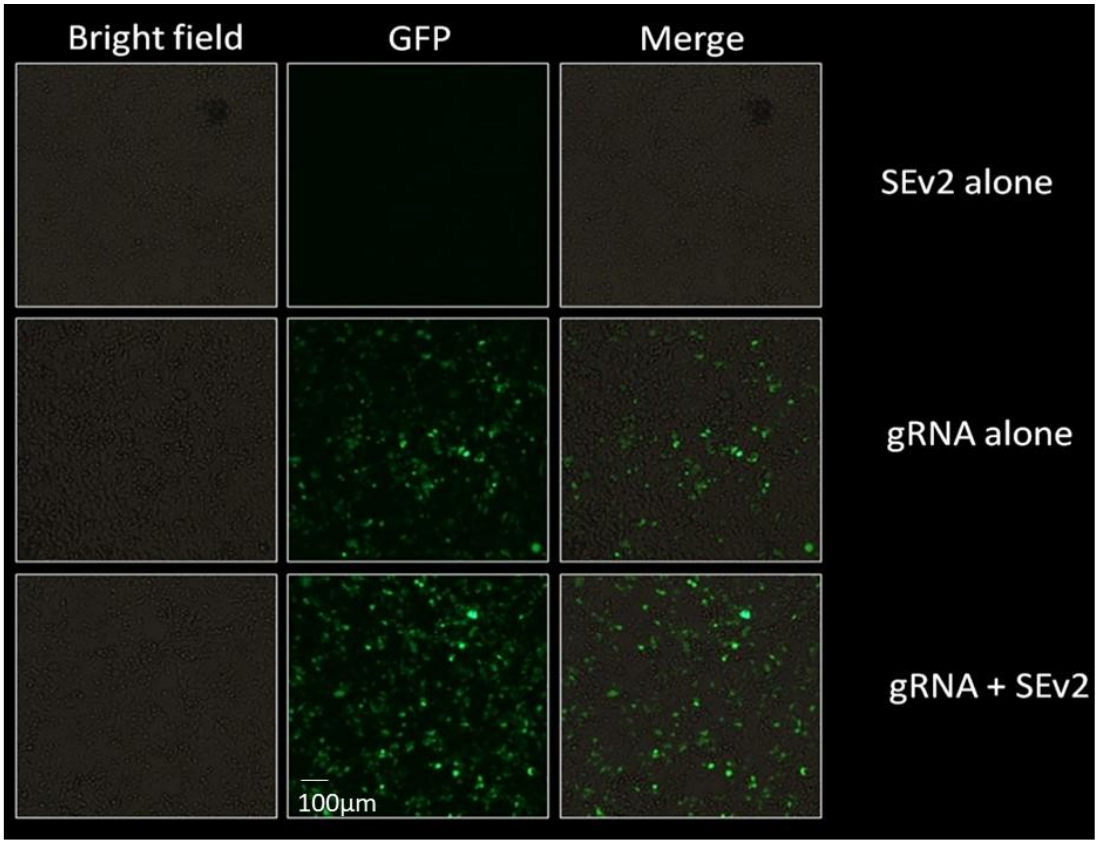


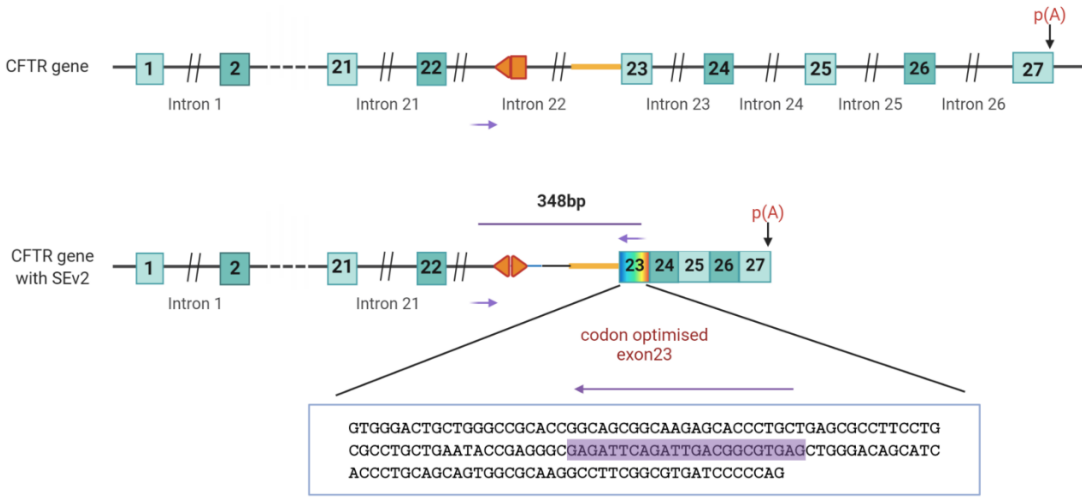
Figure 5.12: HEK293cells transfected with either pcDNA empty plasmid as a negative control (data not shown), SEv2 alone, Cas9/gRNA alone or both Cas9/gRNA + SEv2 together. Twenty-four hours post-transfection, the cells were observed in the microscope to assess the transfection efficiency. For each condition, the bright field images are shown to display the quantity of cells present and the GFP images are shown to display the amount of cells transfected. Magnification 20x. Scale bar = 100µm.
This result is coming from for different transfection experiments (n=4).

Figure 5.12 – Microscopy results of the SEv2 experiment in HEK293 cells

Visually, when comparing the bright field images with the GFP images, the transfection efficiency can be estimated to be around 50%, which is similar to slightly lower to what was observed previously for SEv1.

Eight days post-transfection, the DNA was extracted, and a PCR of the junction was done using a set of primers targeting intron 22 before the cut (NGS 5' FW) and the optimised sequence of exon 23 specifically (exon 23 codon opt REV; Figure 5.10C). Using this pair of primers, only the DNA containing the superexon will display a band. The results of the PCR are shown in figure 5.13B.

A)



B)

n=3

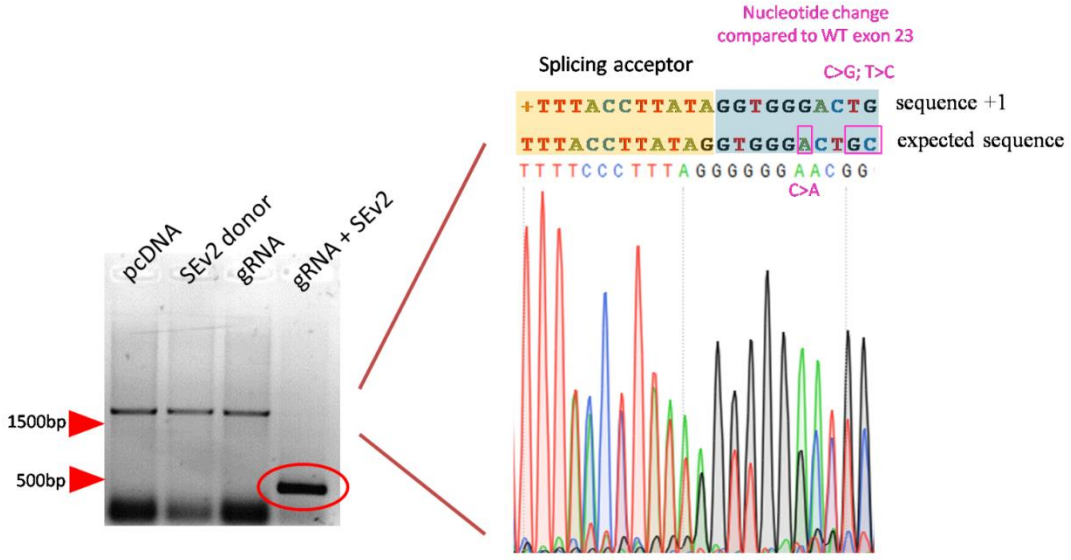


Figure 5.13: Evidence of SEv2 integration in HEK293 cells. The cells transfected in section 5.3.2 were used for DNA extraction 8 days post-transfection. A PCR was done around the 5' junction to assess SEv2 integration. A) Schematic representation of the WT CFTR gene and a CFTR gene with SEv2 integration through HITI. The sequence of the optimised exon 23 and the localisation of the primers are shown in the figure. B) PCR results showing a band of expected size for SEv2 integration, only in the sample co-transfected with both Cas9/gRNA and SEv2. The band of interest was sent for Sanger sequencing to confirm the 5' junction. The Sanger sequencing chromatogram is shown on the right and shows presence of the superexon with the codon optimised exon 23. A mix of two sequences can be observed, the expected sequence as well as a (+1) insertion sequence.
This result is coming from 3 different transfection experiments (n=3).
Figure 5.13 - Evidence of SEv2 integration in HEK293 cells

The result of the PCR showed the expected 348bp band only when both the gRNA/Cas9 and SEv2 plasmids were co-transfected (Figure 5.13B). The presence of this band strongly indicates the integration of the superexon into the intron 22 of the *CFTR* gene. For further confirmation, the band was sent for Sanger sequencing (Figure 5.13B).

Due to the primers design, the forward primer sequence was very close to the junction between the genome and the donor. Usually, after Sanger sequencing it happens that the few first nucleotides are missing and not showing a clear sequence. Therefore, when the amplicons were sequenced, the beginning of the sequence, which corresponded to the junction, was not available to analysis.

Nevertheless, it was still possible to see the presence of the superexon inside the sequence. In the Sanger data file, a mix of two main sequences appears to be present at about the same quantity in the mix of amplicons. Those sequences indicated an integrated superexon sequence without indels (about 40%) and a superexon integrated with a +1 insertion (about 40%). A presence of a mix of other indels also seemed to be present at a lower frequency (about 20%). The data are in accordance with section 5.2, showing integration of the superexon into the *CFTR* gene, and demonstrated a (+1) insertion as the major indel pattern.

5.3.3 - INTEGRATION OF SEV2 INTO THE *CFTR* GENE OF 16HBE14o$^-$ CELLS

Having shown that the SEv2 can be integrated in the genome, the next step was to estimate the integration efficiency and confirm that the integration of the superexon leads to a stable mRNA and functional CFTR protein production.

In order to analyse those parameters, W1282X 16HBE14o$^-$ cells were used. Those cells possess an active *CFTR* promoter and contain the W1282X mutation. Because of the W1282X mutation, the *CFTR* mRNA is not stable, and CFTR protein cannot be detected in these cells by western blot (Figure 3.8 in section 3.5.2; (175)) .

The hypothesis to test here is that the SEv2 donor would integrate in *CFTR* through HITI and correct the W1282X mutation, leading to a stabilised mRNA and a functional protein.

The W1282X 16HBE14o$^-$ cells were transfected with either pcDNA empty plasmid, SEv2 donor plasmid alone, pSpCas9(BB)-2A-GFP (Cas9/gRNA) plasmid alone or Cas9/gRNA + SEv2 plasmids together.

After 24 hours incubation, the cells were observed in a Leica DMI 3000 B fluorescent microscope, using the Leica Application Suite software (LAS v4.1) to capture images, in order to estimate the transfection efficiency (Figure 5.14).

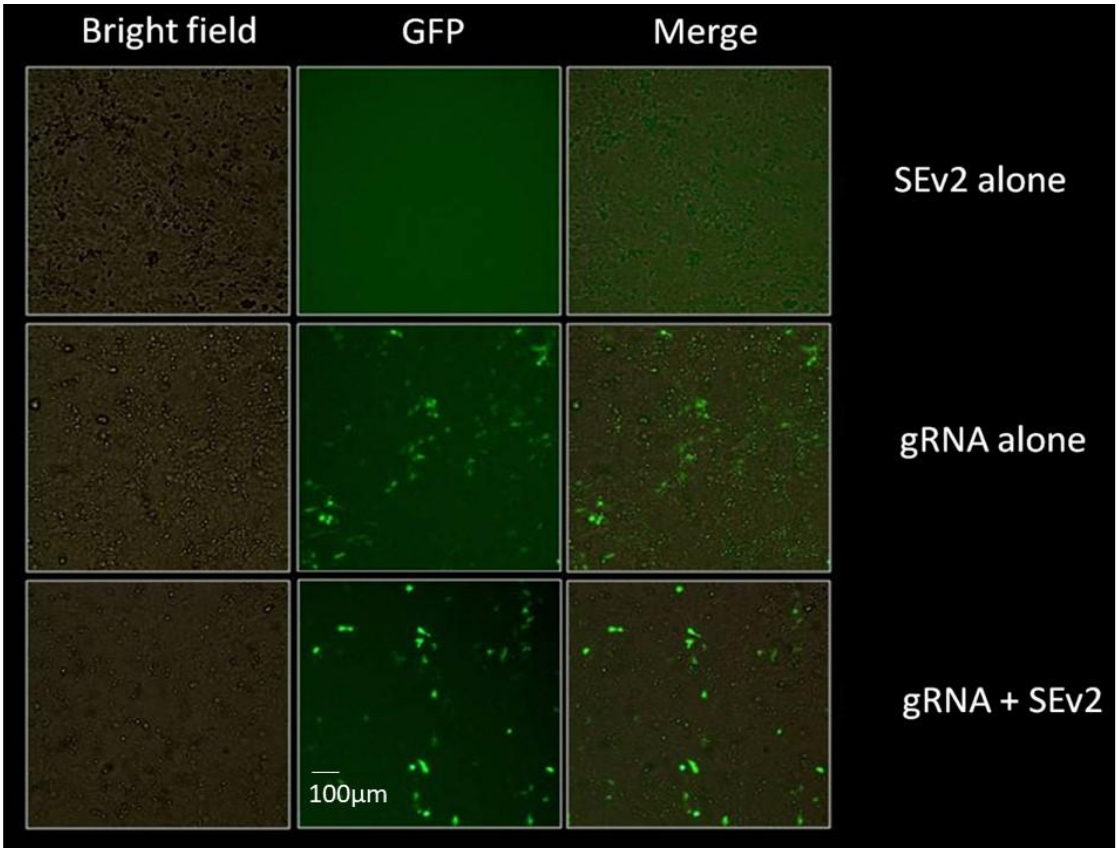


Figure 5.14: W1282X 16HBE14o⁻ cells transfected with either pcDNA empty plasmid as a negative control (data not shown), SEv2 alone, Cas9/gRNA alone or both Cas9/gRNA + SEv2 together. Twenty-four hours post-transfection, the cells were observed in the microscope to assess the transfection efficiency. For each condition, the bright field images are shown to display the quantity of cells present and the GFP images are shown to display the amount of cells transfected.
Magnification 20x. Scale bar = 100µm.
This result is coming from a unique experiment (n=1).

Figure 5.14 – Microscopy results of the SEv2 experiment in W1282X 16HBE14o⁻ cells

Visually, when comparing the bright field images with the GFP images, it can be noticed that the transfection efficiency was about 10 times lower in 16HBE14o⁻ cells compared to HEK293 cells. Because of the low transfection efficiency, sorting of the GFP-positive cells was necessary in order to enrich the cell population for transfected cells, in order to facilitate the identification of corrected cells.

Forty-eight hours post-transfection, the cells were analysed through a FACS cell sorter, and the GFP-positive cells were plated. The cell sorter analysis showed a transfection efficiency of about 5%, consistent with the efficiency estimated in the microscopy results (data not shown).

Two options were available for the seeding of the cells after FACS sorting: YIELD and PURE.

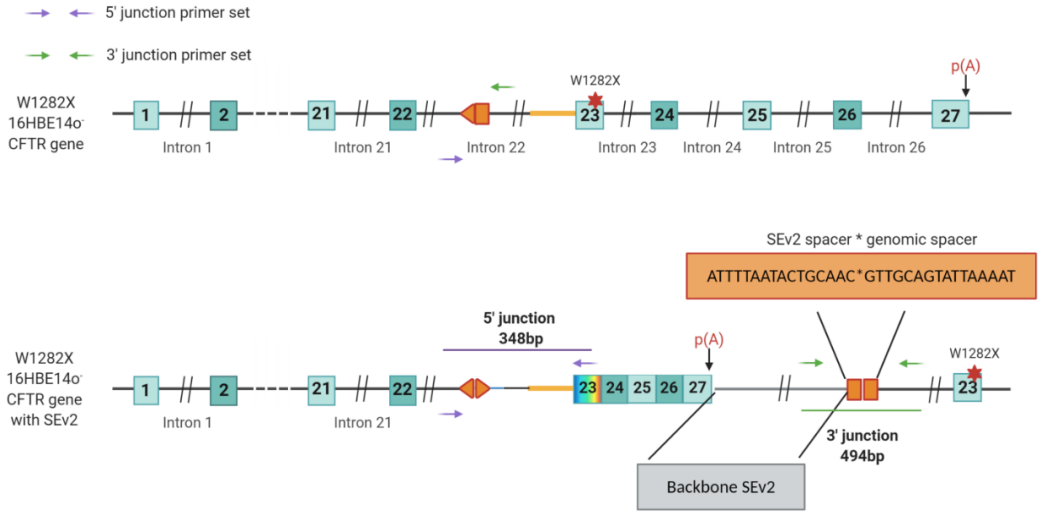
The YIELD option consisted of getting as many GFP-positive cells as possible. In this case, if the machine analyses a cell and is not 100% certain that the cell is really GFP-positive, it will still sort this cell as a green cell. With this parameter, it was possible to have more cells sorted, but the purity of transfected cells could be lower.

The PURE option was to programme the machine to sort only the GFP-positive cells that are the brightest, and therefore, even though there will be less cells selected, all the cells will be transfected cells.

Both options were used for further experiments.

Nineteen days post-transfection, the genomic DNA of sorted and non-sorted cells was extracted and analysed by PCR using two different primer sets to check both for the 5' junction and 3' junction. For the 5' junction, the same primer set as for the section 5.3.2 was used. For the 3' junction, the primers used were targeting the plasmid backbone from the donor (backbone FW SEv2), and the genomic intron 22 after the gRNA recognition site (end intron 22 REV). Theoretically, using those set of primers, only the *CFTR* gene with the superexon integrated will display a band on the electrophoresis gel (Figure 5.15A).

A)



B)

n=1

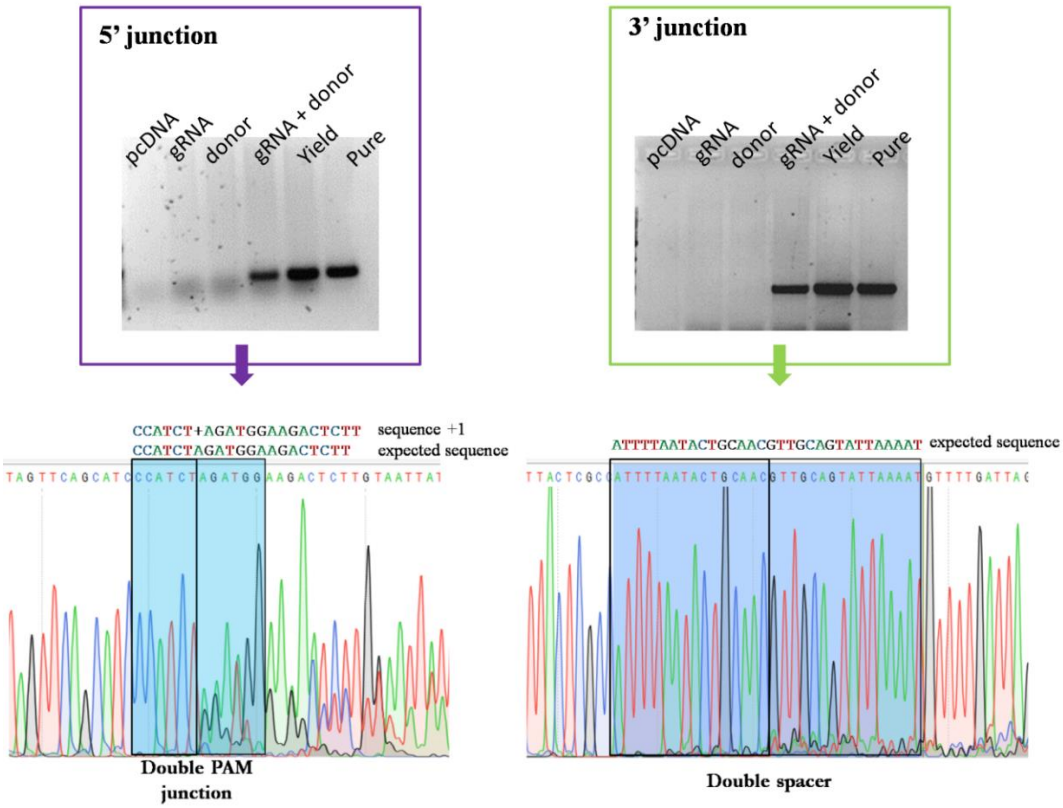


Figure 5.15: Evidence of SEv2 integration in W1282X 16HBE14o⁻ cells. The cells transfected in section 5.3.3 as well as the sorted GFP⁺ cells were used for DNA extraction 19 days post-transfection. A PCR was done around the 5' junction and the 3' junction to assess SEv2 integration. A) Graphic representation of the unedited and edited genome, with localisation of the primers. B) PCR results showing a band of expected size for SEv2 integration, only in the samples co-transfected with both Cas9/gRNA and SEv2, sorted or not. The bands of interest were sent for Sanger sequencing to confirm the junctions. The Sanger sequencing chromatogram is shown below each respective PCR. The sequencing demonstrates presence of the superexon with HITI integration being confirmed by the junction of the two PAMs and the two spacers from the genome and donor.
This result is coming from a unique experiment (n=1).

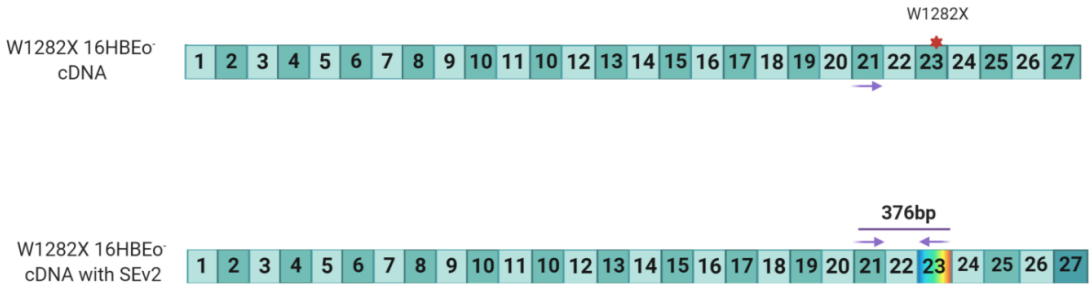
Figure 5.15 - Evidence of SEv2 integration in W1282X 16HBE14o⁻ cells

PCR results on figure 5.15B displayed a band only for the samples where both the gRNA/Cas9 and the SEv2 plasmids were co-transfected. The integration of the superexon into the *CFTR* gene was demonstrated by Sanger sequencing of the bands of interest. HITI was confirmed by the presence of the two PAMs and the two spacers sequences joined together. Interestingly, indels at the 5' junction were high (about 50%), with the majority of them being a (+1) insertion, in accordance with previous results. However, at the 3' junction, only a small background of indels was observed (~5%).

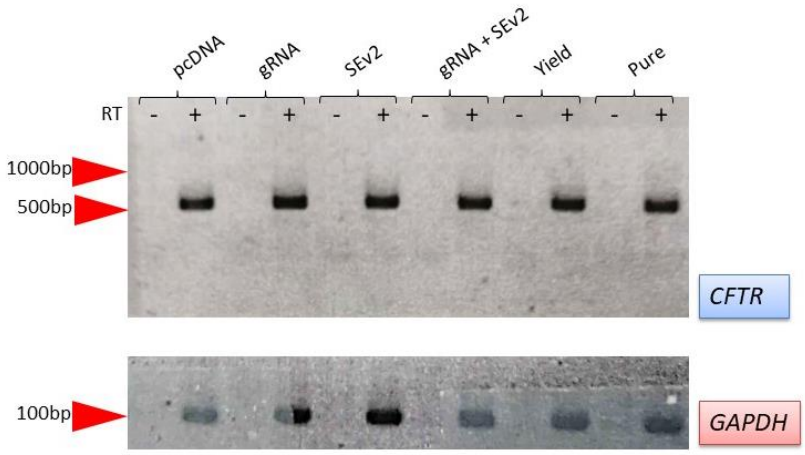
5.3.4 - EXPRESSION OF SEV2-INTEGRATED *CFTR*

The confirmation of the superexon integration in 16HBE14o$^-$ cells, which are CFTR-expressing cells, led to the next goal of this study, which was to determine if the superexon integration led to CFTR expression. For this purpose, one-month post-transfection, the RNA of the cells was extracted, and an RT-PCR was performed to confirm presence of the superexon in the cDNA. The set of primers used was targeting the exon 21 of *CFTR* (CFTR exon 21 FW), and the optimised exon 23 from the superexon (exon 23 codon opt REV). With this set of primers, only cDNA samples containing the integration should display a 376bp band (Figure 5.16A). GAPDH primers were used as controls (data not shown).

A)



B)



C)

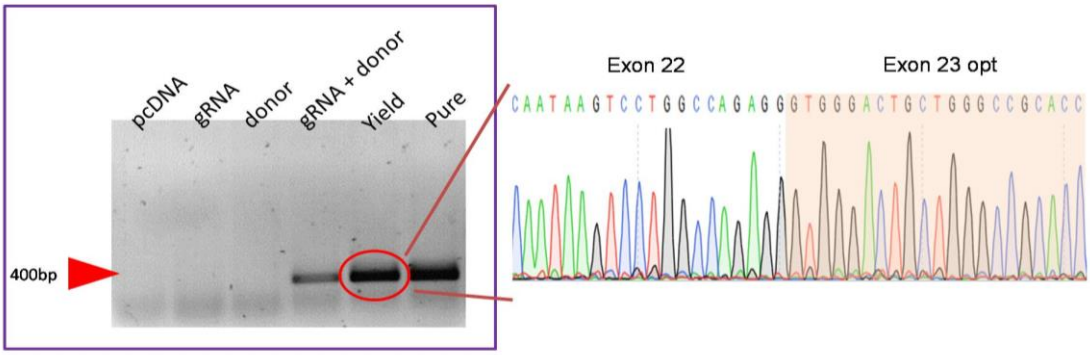


Figure 5.16: Confirmation of superexon-integrated CFTR expression in W1282X mutant 16HBE14o- cells. One-month post-transfection, the mRNA was extracted from the sorted and non-sorted cells. A RT-PCR was performed using a reverse primer specific for the SEv2 construct. A) Graphic representation of the unedited and edited cDNA, with localisation of the primers. B) RT-PCR results using the CFTR exon 21 FW and junction 24-25 CFTR REV primers to confirm mRNA purity. The top gel corresponds to CFTR amplification and the lower gel corresponds to a GAPDH control amplification (GAPDH FW and GAPDH REV). For each condition, a control without reverse transcriptase was performed to assess the purity of the mRNA. C) PCR results showing a band of expected size for SEv2 integration only in conditions transfected with both Cas9/gRNA and SEv2 together, sorted or not. The bands were sent for Sanger sequencing and the SEv2 integration was demonstrated by the presence of the exon 23 optimised, in the cDNA.
This result is coming from a unique experiment (n=1).

Figure 5.16 - Evidence of SEv2 expression in W1282X 16HBE14o- cells

Figure 5.16B shows the PCR results for the cDNA samples. The electrophoresis gel shows presence of the expected 376bp band only when both the Cas9/gRNA and the SEv2 plasmids were co-transfected, while the other conditions did not display any bands. Those bands were sequenced and confirmation of SEv2 integration was shown by the presence of the codon optimised exon 23 sequence, which was only present in the SEv2 plasmid (Figure 5.16B).
To conclude, those results confirmed the expression of the superexon-integrated *CFTR* gene, in W1282X 16HBE14o- cells, with W1282X correction.

5.3.5 - EFFICIENCY OF INTEGRATION AND EXPRESSION OF SEV2 IN 16HBE14o- CELLS

The integration of the superexon in 16HBE14o- cells and its presence in *CFTR* transcripts were demonstrated. However, no indication of integration efficiency could be extracted from the previous experiments. To analyse the integration efficiency in genomic DNA, and the frequency of corrected mRNA, **N**ext **G**eneration **S**equencing (NGS) was performed as stated in section 2.6.

The amplicons used for NGS were coming from two distinct PCRs performed on the Cas9/gRNA + SEv2 PURE sample. The first PCR used a primer targeting a sequence upstream the gRNA cutting site (NGS 5' FW) and the primer sequence designed especially for NGS analysis, between the gRNA recognition site and the splicing acceptor (NGS 5' REV; Figure 5.10). The second PCR used primers targeting exon 21 (CFTR exon 21 FW) and the junction between exon 24 and exon 25 (junction 24-25 CFTR REV).

After NGS, the sequencing files were analysed using the QIAGEN CLC Genomics Workbench software. Preliminary results showed 2736/46555 reads (5.8%) of genomic DNA amplicons contained the superexon sequence. Concerning cDNA amplicons, 2603/33983 reads (7.6%) contained the superexon sequence.

Those results indicate a good efficiency of integration, similar to what was previously observed for superexon experiments using HDR (~10% after selection)(200). Noticeably, cDNA results showed a slight increased, but not as drastic as what was observed for Cas12a and base editing experiments.

Importantly, these NGS analyses are preliminary, with further more detailed analysis on-going. Indeed, the NGS analyses were performed by me, without any previous training. Therefore, another analysis after training will need to be performed in order to confirm or refute those data.

5.4 CONCLUSION AND DISCUSSION

To conclude, this chapter showed that it was possible to integrate a superexon inside the *CFTR* gene using HITI. This technique is powerful due to the use of only one donor to correct most of CF mutations, if a superexon from exon 2 to 27 could be designed. In this chapter, the efficiency of correction was shown to be around 6%, which is comparable to the 10% superexon integration, after selection, seen in previous data (200). cDNA data also showed that the superexon could be expressed. However, no significant mRNA stabilisation from the deep sequencing data were observed compared to Cas12a and Cas9 HDR (section 3.4) or Cas9 ABE (section 4.2) experiments.

One possibility to explain these results could be that, during mRNA splicing, the superexon could be skipped due to a weak splicing acceptor. Indeed, it could be possible that the cells, sometimes, do not recognise the superexon splicing acceptor but recognise the endogenous

exon 23 splicing acceptor. This would lead to the superexon being spliced out. However, the superexon was designed to contain the exact same splicing acceptor sequence as the endogenous one. Therefore, it does not seem like the splicing acceptor sequence is responsible for the superexon skipping.

However, it could be possible that some sequences before the 102bp of the end of intron 22, contain a splicing signal that would have been displaced or removed when the superexon was integrated.

Moreover, a weak polyadenylation site could also lead to mRNA being unstable and degraded. Indeed, if the polyadenylation is not recognised, the transcription will continue until the endogenous polyadenylation site, leading to an aberrant mRNA.

Another option to explain why the corrected RNA was not substantially stabilised is that any of the introns 23 to 26, or the 3' UTR sequence, could play a role in the downregulation of the expression of the superexon-integrated *CFTR*. Indeed, all those introns and most of the sequence from the 3'UTR were removed in the superexon design. It is now known that the *CFTR* gene contains miRNA sites, even though those miRNAs are not well-established yet. For example, a paper from 2011 described about 13 to 15 miRNA localised in the 3' UTR of the gene (201). Those miRNAs are expressed differently regarding of the cell type and have been shown to have a downregulating effect on the gene expression. It is believed that those miRNAs could play a role in the cell-specificity of *CFTR* expression. Prediction tools to search for miRNA target sequences predicted between 255 and 496 putative *CFTR* targeting miRNAs (TargetScan; (202, 203)). Moreover, the length of the *CFTR* 3'UTR is more than two times longer (1.5kb) than the average for human genes (740bp). All this show that *CFTR* regulation could be very complicated, and cutting out non-coding DNA could potentially negatively affect those regulations. More characterisations need to be done on *CFTR* regulation and non-coding RNA, before considering bringing a *CFTR* superexon constructs intro clinic.

Those sequences, and mostly the 3'UTR region of genes, have been shown previously to contain important sequences such as microRNA target sequences (204). The assumption here is that those missing miRNAs target sites could either inhibit the superexon splicing or downregulate the expression of the superexon-integrated *CFTR*.

Overall, this technique seemed efficient and promising to correct multiple CF mutations. To focus on the safety data, the DSB in intron 22 created indels in that region, and those indels were estimated to be around 50%. Because HITI lay on NHEJ repair pathway, the indels were expected. However, the position of those indels inside the intron 22 is thought to be less harmful than other techniques creating indels inside an exonic sequence. Nevertheless, it will be important to check for the protein expression and function in many corrected clonal cell populations, in order to first, confirm that superexon integration leads to a functional protein, and second, to validate the safety of indels in this region of the intron.

Another advantage of this technique is that, even if the gRNA recognition site was disrupted by indels, another subsequent donor administration could be done by using another gRNA recognition site downstream the first cut site. The guide sequence could be easily replaced in the SEv2 plasmid. This way, cells that have not received the superexon would have this intronic region and could undergo HITI. This second administration could, in theory, increase the percentage of correction.

CHAPTER 7 - OVERALL CONCLUSION AND DISCUSSION

Cystic fibrosis is a fatal genetic disease caused by CF-causing variants in the *CFTR* gene. There is actually no cure for CF, but treatments are available to ease the symptoms. CFTR modulators are molecules that improve the processing and function of the CFTR protein, to treat some patients with specific CF-causing variants (section 1.4). However, the cost for those treatments are expensive, cumbersome and don't benefit 100% of patients. Indeed, some CF-causing variants, such as variants from the class I, express few to no proteins and therefore, patients with those variants cannot benefit from any treatment targeting the CFTR protein (section 1.3.3). The variant W1282X is one of them. The W1282X mutation is the 6th most common CF-causing mutation, concerning about 1.2% of CF patients, moreover, it is the 2nd most common class I variant. Since the discovery of the *CFTR* gene in 1989, multiple clinical trials for CFTR cDNA addition have been performed, unsuccessfully (section 1.5). However, since the discovery of programmable nucleases, for genome editing, new hopes for CF gene therapy emerged.

The overall goal of this project was to compare four different techniques to correct the W1282X mutation, either in a sequence specific manner, using HDR (aim 1) and base editing (aim 2), or as a superexon to correct this mutation and all the ones downstream. Aim 3 has for purpose to integrate a superexon using HITI, while the objective 4 has for purpose to replace exons with a superexon using an HDR donor template. The objective was to determine if there is one technique that is optimal for CF purpose.

7.1 CONCLUSION OF THE WORK

7.1.1 OBJECTIVE 1

In the first objective, using the HDR pathway, SpCas9 and AsCas12a were compared. AsCas12a was interesting since this protein was shown to be more specific (i.e.- with fewer off-target effects than SpCas9), making this protein potentially safer (112).

However, it was shown in chapter 3 that SpCas9 was more efficient than AsCas12a, with 18% DNA correction vs 7% for AsCas12a, and 60% RNA correction vs 50% for AsCas12a. Western blot analyses for Cas12a showed that DNA correction of at least one allele was responsible for protein expression.

However, it has not yet been possible to generate functional data from these corrected cells, and no protein analyses have yet been performed for SpCas9 by the other PhD student who did the SpCas9 study. Interestingly, in table 3.1, which shows sequencing analyses of 42 AsCas12a-generated clonal cells, it was found that 3/42 clonal cells (~7%) were homozygous for W1282X correction, and 9/42 clonal cells (~21%) had at least one allele with the W1282X correction. It is known that carriers of a CF-causing mutation are healthy. Therefore, it would be interesting to know if the 21% cells being at least heterozygous for W1282X correction would be enough to compensate for the lack of CFTR from the other cells.

The main disadvantage of the HDR technique, displayed in chapter 3, was that non-corrected alleles contained indels in the exonic sequence, potentially disrupting the *CFTR* gene. Indeed, data analysis from the 42 AsCas12a-generated clonal cells, showed that 32/42 (72%) of the studied cells contained indels on at least one allele, and 45/84 alleles (~54%) contained indels (Table 3.1). This result showed that most of the *CFTR* alleles were disrupted after the experiment. Considering that cells heterozygous for indels and W1282X correction should have their chloride transport functional, there should be 26/42 non-functional cells (~62%) which would contain indels in at least one of their alleles and potentially have their gRNA recognition site disrupted.

Disruption in the gRNA recognition sites could prevent a potential second gene editing treatment, if the first one was not enough to restore CFTR function in the epithelium. For that purpose, AsCas12a could have an advantage over Cas9, since the DSB occurs at the end of the gRNA recognition site. Indeed, in section 1.8.2d, it was described that Cas12a is sensitive to mismatches up to position 18, and the last few nucleotides of Cas12a gRNA

recognition sites were not essential for the nuclease specificity. Short indels around the end of the recognition site should still allow Cas12a to cut the same site. However, DSB created by SpCas9 are localised close to the PAM, so short indels would most likely disrupt completely the recognition site.

Moreover, in the case where a drug such as a read-through, becomes available for W1282X, the 54% alleles that are disrupted would not be responsive to the drug.

This is a significant limitation in HDR technique possibly preventing its translation in the clinic, if the level of indels occurs at a similar frequency *in vivo*. However, if precisely edited cells can be accurately identified prior an *ex vivo* type of therapy, this would be less of a concern.

Recently, two groups demonstrated very high HDR efficiencies to correct cystic fibrosis mutations (152, 208), using a technique described as CRISPR RNP electroporation and AAV donor infection (CRISPR-READI; (209)). Vaidyanathan et al. corrected the F508del mutation using Cas9/sgRNA RNP (152), while Suzuki et al. corrected the same mutation using ZFN mRNA (208). Both groups electroporated the nucleases and transduced the donor using AAV6. After optimisation of the sgRNA, the cell model, the cell passage, the multiplicity of infection (MOI) and the template, the first group showed a frequency of allele correction of 41% ± 6%, with 38% ± 3% indels in upper airway basal cells (152). These results are about two times higher than the results demonstrated in the chapter 3 of this thesis. Suzuki et al. used airway basal cells, and, with their best donor template, which was a 2kb donor spanning the exon 11 of the *CFTR* gene, they got a frequency of editing of 31% ± 4%, with about 20% indels (208). Those results indicate that further optimisation could be possible in order to increase efficiency of HDR.

7.1.2 OBJECTIVE 2

In the second objective, the base editing technique, described in chapter 4, also allowed correction of single mutations. Base editing was about as efficient as Cas9 HDR with 24% DNA correction and 50% RNA correction. The advantage compared to the technique in aim 1 was that no indels were observed within the limit of detection of the Sanger sequencing (Figure 4.4).

However, the main disadvantage is that the editing could lead to "by-stander" mutagenesis of other nucleotides in the editing window, which can be deaminated by the base editor being used to correct the target nucleotide. That by-stander mutation could potentially be a disease-causing mutation. Moreover, a current limitation to base editing is that only C>T (G>A) mutations and A>G (T>C) mutations can be corrected by cytidine base editors and adenine base editors respectively. Conversely, HDR can correct any type of mutation comprising all substitutions, as well as small insertions and deletions, many of which cause a frameshift.

Recently a paper was published, using adenine base editing to correct the W1282X mutation in the *CFTR* gene of 16HBE14o$^-$ cells (210). In this paper, the group used a codon optimised ABE called RA6.3, which had showed an improved efficiency in HEK293T cells. However, like in our hands, plasmids were poorly expressed in 16HBE14o$^-$ cells, but electroporation of mRNA expressed well. They electroporated a uridine depleted RA6.3 adenine base editor mRNA, containing 5-methoxyuridine modifications on the remaining uridines, with a sgRNA, containing 2'-O-methyl 3'-phosphorothioate modifications, to improve the stabilisation of both RNAs *in vivo* (210). After electroporation, the group showed base conversion of the W1282X 5'-TGA-3' to 5'-TGG-3' at a frequency of 26.4% ± 7.4% (210). The editing window was from the 4th nucleotide until the 9th, the first nucleotide being the furthest from the 5'-NGG-3' PAM. In that editing window was two adenines. The adenine from the W1282X mutation was the 9th nucleotide (A_9) while the bystander adenine was placed at the 5th position (A_5; see figure 4.1A for sequence). The group showed that the A_5 was edited at a frequency of 45% ± 5.66%, leading to a change of codon from a glutamine to an arginine (Q1281R) (210). Those results showed that the bystander editing was two times higher compared to the editing of interest.

This converse with the results obtained in chapter 4 with NG-ABEmax, where it was found a similar editing frequency for the both adenines present in the window of editing (A_6 and A_7; Figure 4.1B). However, the group made two clones; one with the W1282X correction and the bystander on the same allele, and one with the W1282X correction and the bystander on the second allele. They showed no effect of the bystander Q1281R on CFTR function (210), demonstrating that this bystander could be safe to introduce.

The next step of the objective 2 will be to identify if the bystander mutation R1283G affects the CFTR function.

Importantly, in 2020, the group of Clevers, showed the use of CBE and ABE to correct four selected CF-organoid samples (190). One of those CF-organoids contained the W1282X mutation. The group used a self-made xABEmax (190) by cloning a codon optimised xCas9 (pLenti-xFNLS-P2A-Puro) described previously by Zafra et al. (211) to the ABEmax plasmid described by Koblan et al. (161). Interestingly, electroporating this self-made plasmid and using the same PAM site used in chapter 4 of this thesis (figure 4.2A), Geurts et al. described high base editing of the A_6 from the W1282X mutation, but no base editing of the A_7. This correction was linked to a functional CFTR repair.

Those results were surprising due to the difference with the results obtained in our hands. Indeed, we tried to correct the W1282X mutation by base editing using the ABE7.10 (Addgene plasmid#102919), the xABEmax (Addgene plasmid#119813) and the xCas9(3.7)-ABE(7.10) (Addgene plasmid#108382), but didn't see any base editing at all (data not shown).

The difference is results could be maybe explained by the extra nuclear localisation signal made by Zafra et al. in their xFLNS plasmid (211), improving the nuclear localisation of the base editor. However, the ABEmax from Koblan et al. had an editing window from A_4 to A_8, so the absence of base editing at position A_7 is surprising. A better characterisation of this self-made plasmid, with characterisation of the editing window would be very interesting. Maybe the cloning of the two constructs could have altered the editing window.

7.1.3 OBJECTIVE 3

The next objectives were to assess the use of a superexon construct as described in chapters 5 and 6, to correct a large number of different mutations, anywhere in the *CFTR* gene, using only one unique construct. Succeeding to integrate a superexon into the human genome would be the ideal therapeutic method for Cystic Fibrosis, since even rarer mutations could be corrected. A potential disadvantage of a superexon construct could be the loss of many or all the introns, which may contain important regulatory sequences such as microRNA target sites (212).

In chapter 5, the integration of the superexon was explored using **H**omology-**I**ndependent **T**argeted **I**ntegration (HITI). This technique takes advantage of the non-homologous end joining pathway, which is the default repair system in every cell. Using HITI, the efficiency was lower compared to HDR and base editing in previous chapters, with about 6% DNA correction and about 8% RNA correction.

Noticeably, this efficiency was obtained after selection of the transfected cells. No protein data have yet been obtained, but it would be interesting to see if this efficiency of correction could lead to a functional improvement in ion transport. Papers from 1995 and 2002 described that a mixture of about 5-10% non-CF cells could potentially correct chloride abnormality (213) and about 5% of wild-type full length mRNA in every cell should lead to an improvement of CF symptoms (214).

Considering those papers, the use of HITI is close to the values that are expected to be beneficial. However, at the moment, no *in vivo* studies confirmed CF correction, therefore there are no available data confirming the therapeutically beneficial amount of correction. For example, is 6% correction with HITI going to restore chloride transport, and are the 8% of mRNA enough for those 6% only? Thus, the superexon integration using HITI is close to what is expected to be beneficial, but it might be necessary to improve the efficiency to expect a beneficial therapeutic effect. Moreover, the non-stabilisation of the W1282X-corrected mRNA was unexpected and would need more characterisation.

An advantage of HITI, which could be also a disadvantage, is that a lot of indels (up to 50%) could be observed in the intronic sequence (Figures 5.13 and 5.15). The fact that the indels are intronic should limit the potentially harmful effect, compared to the indels observed in the exonic sequence, for HDR, in chapter 3.

However, more in-depth studies will need to be performed after a superexon design, to confirm the safety of those indels. For example, some intronic regulation sites could be damaged, or formation of new splicing site could emerge.

The use of a superexon to correct genetic diseases was first described *in vivo*, in a humanised mouse model of Haemophilia B, by Li et al. (170). They designed a donor containing a splicing acceptor, a superexon from exon 2 to 8 of the *hF9* gene, and a polyadenylation site, all of it flanked with homology arms for HDR. The group designed ZFNs cutting in the first intron. They transduced both the ZFN and the donor using AAV8 vectors. It was described an editing frequency of 1-3%, leading to clinically significant correction of the coagulation defect (170). Moreover, the group described that this AAV8 transduction was well tolerated in mice (170). Those results were encouraging to persevere with a superexon construct for other diseases.

As described in section 1.12, Bednarski et al. used a superexon construct (exons 11 to 27) with HDR, in the context of cystic fibrosis, in CFBE41o$^-$ cells (171). They designed a donor containing *CFTR* exons 11 to 27, with a splicing acceptor upstream and a polyadenylation site downstream, all of it flanked with homology arms for HDR. They designed a ZFN pair that cut inside the exon 11 of *CFTR* and transfected both the ZFN plasmid and donor using Lipofectamine. The group showed successful insertion of the superexon by HDR. Moreover, the superexon integration showed restoration of CFTR function in a F508del CFBE41o$^-$ cell line. However, the integration efficiency using HDR was very low, with 10% integration after a puromycin selection (171).

More recently, Suzuki et al. showed in 2020 the very efficient integration of a superexon from exons 9 to 27, into the intron 22 of the *CFTR* gene (208). The construct also contained homology arms flanking a splicing acceptor, a superexon from exon 9 to 27, and a polyadenylation site. The group electroporated ZFN mRNA and transduced the donor using AAV6, into airway basal cells containing the G542X/R785X mutations. All of these led to an editing frequency of 61.8% ± 6%, and restoration of mature CFTR protein expression and function (208). Those results are encouraging that efficient HDR can be performed to integrated large DNA fragment such as a superexon.

Concerning HITI, very few works have been published using this technique. A publication *in vivo*, has been published in 2020 and showed a knock-in of a 3kb DNA construct composed of the *RS1* gene and a *GFP* cassette (RS1/GFP), in the *Rosa26* site of mice retina's cells (215). This *RS1* gene, when mutated, is responsible for the X-linked juvenile retinoschisis, which impairs vision in males. As a DNA delivery, supramolecular nanoparticles were injected in the eye of the mice (215). GFP presence could be observed on retinal surfaces and it lasted for at least 30 days, until when the mice were euthanised for further characterisation. This showed persistence of GFP expression after HITI integration. DNA extraction studies showed the PAMs and spacers junctions at each end (215), like shown in chapter 5 (Figure 5.14B). Those results are encouraging for the use of HITI for a persistent gene expression in non-dividing cells.

7.1.4 OBJECTIVE 4

Finally, in chapter 6, the final aim of the project was to evaluate the use of a superexon, to correct the W1282X mutation through HDR, to try and achieve a higher editing efficiency than what was observed previously by the group of Cathomen in 2016 (200).

In objective 4 of this thesis, a proof of principle for the feasibility of superexon gene replacement using HDR was performed, replacing the whole exon 23 and some shorter flanking regions (~390bp total). Usually, HDR is used to correct short mutations, with a DSB close to the mutation for greater editing efficiency, within 40bp for ssODN donor (216), or within 200bp for plasmid donor (217). A study from 2015 succeeded to insert more than 2kb template into iPS cells using homologous recombination (HR), by replacing two exons in the human *THY* gene with the same exons from the mouse *THY* gene, at a high efficiency (~25%) without selection. Those two sequences shared less than the 75% identities required for HR (205).

In chapter 6, sequences of the exon 23 (156bp) as well as 230bp of the end of intron 22, were modified in the donor template, so there would be no more than 7bp identity with the endogenous sequence (Figure 6.4B). Transfection of both Cas9/gRNA and the donor template in HEK293 cells and HEK Flp-In 293 EMG i21-i24 W1282X cells showed successful replacement of the whole exon 23, into the *CFTR* gene, through homologous recombination (Figure 6.6).

However, the efficiency of insertion couldn't be assessed in the HEK Flp-In 293 EMG i21-i24 W1282X cells, due to the long sequence homology between the CFTR EMG and the endogenous *CFTR* gene, preventing deep sequencing through Amplicon-EZ sequencing (GENEWIZ). However, the mRNA correction could be analysed through the same technique. Results showed low efficiency of about 0.02% without selection. Analyses through Sanger sequencing seemed to indicate that DNA correction efficiency could be less than 5%, which was the limit of detection of the sequencing chromatogram obtained. Because no preselection of the cells was performed, the technique could not be directly compared with the results obtained from Cathomen's lab in 2016 (200).

Future studies and characterisation will need to be carried out. First, selection of transfected cells might be necessary to achieve values of correction that are easier to analyse. However, because the transfection efficiency in HEK Flp-In 293 EMG i21-i24 W1282X cells seemed good with about 50% GFP-positive cells observed in the microscope (figure 6.5), the efficiency would still be expected to be low compared to what was observed for HITI. Second, DNA correction efficiency needs to be assessed in order to compare them with the RNA correction efficiency. DNA correction efficiency could be assessed using HEK293, however, because those cells don't express CFTR, no correlation with the RNA would be possible. Another possibility would be to analyse the exon 23 optimised insertion in the HEK Flp-In 293 EMG i21-i24 W1282X cells, using a different deep sequencing technique that would allow sequencing of a bigger sequence. The relationship between DNA and RNA will give information to know if the low level of RNA editing efficiency could be due to a low DNA correction efficiency or a low splicing.

Ways to improve the efficiency of this technique could be to increase the length of the homology arms, as shown in the paper from Byrne et al. where they found the best HR-mediated gene insertion efficiency with homology arms of 2kb (205). The homology arms used in chapter 6 were only about 130bp long.

Moreover, a paper from 2018 described a technique called targeted integration with linearized dsDNA-CRISPR (tild-CRISPR), where a high knock in efficiency in mouse embryos (33%) was observed when the donor contained 800bp homology arms and was double stranded and linearised (218). Even though this technique was described for mouse embryos, it would be interesting to try it on HEK293 cell lines, to assess if the HDR efficiency could be increased.

The plasmid used in chapter 6 was not linearised and the next step could be to linearise it and compare the HDR efficiencies. This linearisation could be done by adding a gRNA target site for a DSB inside the cell, or by making a PCR-amplified donor or enzyme-cut before transfection.

As for the HITI technique, the indels are located in the intronic sequence and will need deep characterisation for safety. Nevertheless, this technique of HDR in chapter 6 is an advantage for safety compared to HDR for single mutation, like the one described in chapter 3, for which exonic indels can induce gene disruption.

The advantage of this technique over HITI is that only the non-corrected alleles would contain indels, while the corrected alleles wouldn't contain any indels. Whereas, with HITI, potentially every allele cut by Cas9, with or without donor integration, could have indels.

Overall, correcting a single mutation was proved to be more efficient, while inserting a superexon was proved to be still challenging due to low efficiency. Concerning safety, base editing, and the superexon techniques seemed to be the safest.

7.1.5 SUMMARY

The figure 7.1 summarises the different advantages and limitations for each technique described.

	Cas9 HDR for single mutations	Cas12a HDR for single mutations	Base editing	HITI for Superexon	HDR for Superexon
Efficiencies	DNA : 18% RNA : 60% Protein : N/A High efficiency	DNA : 8% RNA : 50% Protein : expression Medium efficiency	DNA : 24% RNA : 50% Protein : N/A High efficiency	DNA : 6% RNA : 8% Protein : N/A Low efficiency	DNA : N/A RNA : 0.02% Protein : N/A Very low efficiency
Advantages	High efficiency	Medium efficiency Potentially lower off-targets than Cas9	High efficiency No detectable indels	Can correct any mutation Advantageous for rare mutations Presence of indels in intronic sequences	Can correct any mutation Advantageous for rare mutations Presence of indels in intronic sequences
Limitations	Indels in exonic sequences →Gene disruption Correct only one mutation → Disadvantageous for rare mutations	Indels in exonic sequences →Gene disruption Correct only one mutation → Disadvantageous for rare mutations	By-stander mutations in the editing window Can correct only C>T (G>A) or A>G (T>C) mutations	Low efficiency Presence of indels in intronic sequence	Very low efficiency Presence of indels in intronic sequence
Off-targets	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**

Table 7.1: Table summarising pros and cons for each technique studied in this project. The green cases represent great advantages, orange cases represent slight limitations and red cases represent big limitations.

Table 7.1 - Summary of the pros and cons for each genome editing technique studied

7.2 DISCUSSION

7.2.1 CELL TARGET

Noticeably, it can be seen from the figure 7.1 that there is no technique that is perfect to do genome editing in CF. Moreover, the choice of technique to be used could be influenced by the cell targets. At the moment, the cells to target using genome editing in CF are not well defined. Recently, a new subtype of cells called ionocytes, were shown to express more than 50% of *CFTR* mRNA (31, 32). Those cells could be a target for CF, however, they compose only up to 1-2% of cells in the airways (31, 32). Therefore, they could be difficult to target specifically.

Interestingly, it could be considered that ionocytes express more CFTR due to a regulation system that is activated or inactivated. It could be assumed that those regulations prevent CFTR expression in other cells. Therefore, studying ionocytes and identifying those regulators, could allow researchers to activate CFTR expression in other cells, compensating for ionocytes CFTR impairment in CF.

Nevertheless, one should question if allowing CFTR expression in other cell types could compensate for the lack of CFTR production from the ionocytes, and if activating CFTR expression in cells that usually do not, could potentially be harmful for CF patients.

In humans, airway cells are not well defined. However, in mice, airway cells are mostly non-dividing cells and their lifespan are between 6 months, in upper airways, and 17 months, in the lungs (219). If most of the airway cells are non-dividing cells, the HDR techniques for both single mutation correction and superexon integration, would be very inefficient. Indeed, it was shown in section 1.9, that HDR is only active during G2/S phases (122).

Importantly, since those cells will eventually die after potentially few years, repeating doses of a gene therapy treatment will be necessary. Indeed, when the cells die, the transgene will be lost, since the cells do not divide into daughter cells containing the transgene.

Another cell target for CF gene editing, could be the basal cells, which are multipotent cells. Using those cells, any of the four techniques described above could be used. However, since basal cells are few and not exposed, targeting specifically those cells *in vivo* seems challenging. Another way to correct those cells, could be using *ex vivo* gene editing, where the cells are extracted from a patient, corrected, and then inserted back in the organism.

Because basal cells have the particularity to differentiate into any other airway cell types, such as ionocytes, the endogenous regulation systems and endogenous CFTR amounts could be conserved. In 2020, the group of Matthew Porteus described *ex vivo CFTR* correction in multipotent airway cells, using HDR and CRISPR/Cas9, with high efficiency of about 20 to 50%. The major drawback for the use of *ex vivo* gene therapy for CF is the lack of successful transplantation strategies of the corrected stem cells (152). Using basal cells, *ex vivo*, any of the four techniques described above could be used.

7.2.2 SAFETY

Concerning the safety of the four techniques studied in this thesis, base editing, as well as HITI and HDR superexon integration, seem to be the safest to use regarding indel profiles. However, concerning off-target safety profiles, no experiments have yet been done. Off-targets can cause severe issues in a cell. Those issues could be chromosomal rearrangements, causing damages at the targeted loci. Moreover, the off-targets could potentially affect genes that can disturb some important physiological functions and cell-signalling (108). Several techniques have been developed and used for off-target detection. First, knowing that the Cas9 protein is more sensitive to mismatches in its seed sequence, and that it can tolerate up to 6 mismatches (section 1.8.1d; (97)), *in silico* tools could be used to predict which genomic sites could be potential off-targets, by analysing the percentage of identity and the location of the mismatches. Once the potential off-target sites are determined, deep sequencing analyses can be done at those sites to detect indels formation. This, as well as other techniques, are well described by Manghwar et al. in their review from 2020 (108).

Moreover, for base editing, papers showed that about half of gene-coding RNA had at least one of their adenine modified, the development of tools such as SECURE-ABE that limits ABE off-targets could be safer to use as an optimisation (168).

7.2.3 OPTIMISATION OF EFFICIENCIES

Concerning the efficiencies, each technique could potentially be studied in greater depth and optimised to improve efficiency. For example, a paper from 2020 showed *CFTR* correction of the F508del mutation using homologous recombination with an efficiency of about 42% of allelic correction and about 38% indels, using RNP electroporation for Cas9/gRNA and AAV6 transduction for the donor (152). Other ways to optimise those techniques could be using a more efficient Cas12a that is now available to increase efficiency of Cas12a HDR (Cas12a Ultra available on IDT; (118)). In addition, a more effective base editor, ABE8, has been recently described for base editing (220). A concern with ABE8 could be that even though the efficiency of adenine deamination at positions A_5 to A_7 is higher, higher deamination efficiencies were also described at positions A_3-A_4 and A_8-A_{10}, potentially increasing the potential number of by-standers products. Moreover, for optimisation of efficiency, it could be suggested that making a superexon donor smaller, for example as a minicircle, could improve efficiency and safety for HITI, compared to a plasmid donor.
Minicircles are supercoiled expression cassettes coming from a bacterial parental plasmid vector and having the majority of their bacterial backbone sequences depleted. Minicircles are minimal vectors with a reduced size, showed to enhance gene delivery and cell viability (197).

7.2.4 CF'S ANIMAL MODELS

After optimisation of those techniques, it will be important to test them in CF animal models before translating them into therapeutics for humans. However, the main limitation in CF is the lack of animal models reproducing the same symptoms as in humans. Different CF animals were designed such as mouse, rat, zebrafish, rabbit, sheep, ferret or pig. Mice and rats do not display any substantial airway infection, and therefore, those models will be difficult to use to study the correction of the CF symptoms. Ferrets and pigs, however, display lot of similarities to CF symptoms in humans, comprising lung disease as well as intestinal and pancreatic diseases. However, those severe diseases led to the death of the animal shortly after birth, preventing long-term studies.

To address this, in 2013, some gut corrected CF pigs were designed and allowed pigs to live longer (221). More recently, *CFTR*$^{G551D/G551D}$ ferrets, responding to VX-770 treatments, were designed (222). When those ferrets were treated *in utero* and after birth with ivacaftor, the multiorgan disease was effectively delayed until the withdrawal of the treatment. Upon withdrawal of VX-770, significant CF disease symptoms appeared within a month. The advantage of this latest ferret model is that CF can be studied at any desired developmental stage, by removing the VX-770 treatment. However, those ferrets have only the G551D mutations, and only gene editing for this mutation can be used with this model. Nevertheless, any technique studied in this thesis could be tested for G551D correction in the *CFTR*$^{G551D/G551D}$ ferrets. Indeed, G551D (c.1652G>A) mutation is a base substitution from a G to a A, changing the codon 5'-GGT-3', which encode a glycine, into the codon 5'-GAT-3', encoding an aspartic acid. Using the adenine base editing the adenine can be changed back into a glycine, restoring the sequence of origin. HDR can also be used to specifically change the adenine into glycine in the ferret genome. Importantly, the superexon23-27 that was designed in this project will not correct G551D mutation, which is located in exon 12. However, a longer superexon comprising exon 11 (which contains the most common mutation F508del) and all the downstream exons, could potentially correct the G551D mutation.

In conclusion, those *CFTR*$^{G551D/G551D}$ ferrets could be a good model to assess the different gene editing techniques studied in this project, in order to further optimise or to validate their safety for a potential use in humans, however, every gRNA or donors will need to be re-designed for the G551D mutation.

Interestingly, the paper from the *CFTR*$^{G551D/G551D}$ ferrets showed that even though 95% of the VX-770 treated kits survived meconium ileus, 63% of the *CFTR*$^{G551D/KO}$ kits also survived (compared to 22% for *CFTR*$^{KO/KO}$ kits; (222)). This shows that it could be possible to design some heterozygous ferrets with the responsive G551D mutation and another mutation, to study the latter. Indeed, if a *CFTR*$^{G551D/W1282X}$ ferret could be designed, it could be possible to use the G551D mutation for the ferret survival, and the W1282X mutation to study and characterise the gene editing techniques designed during this project.

7.2.5 ADVANCEMENT IN DNA DELIVERY VEHICLES

To insert CRISPR/Cas systems and donor into the airway cells, both viral and non-viral vectors can be used. Those systems were described in section 1.5.1. As stated, no vectors to date have proved to be successful in human trials. However, more work is being done to improve viral vectors in order to increase airway apical surface targeting and decrease immune response. The last promising updates on gene therapy in CF, were from lentiviruses and AAV. First, Alton and collaborators described a **S**imian **I**mmunodeficient **V**irus (SIV) pseudotyped with the Fusion (F) and **H**emagglutinin/**N**euraminidase (HN) proteins from SeV. This synthetic lentivirus has been shown to be safe to re-administered, and moreover, could integrate the cDNA into the genome, making the expression permanent (223). However, because integrated-viral vectors have previously been shown to cause insertional mutagenesis (78), an intensive screen for safety needs to be done before using this vector in humans.

Second, some hopes also emerged for AAV vectors, with the discovery of AAV1 and AAV5, which were shown to be more efficient than AAV2 at transducing airway cells (224, 225). Moreover, AAV was extensively used for gene therapy in other diseases due to its moderate adverse effects (226). However, studies showed that upon repeated administration, the efficiency of transduction decrease due to formation of neutralising antibodies (227). Finally, one of the main limitations of AAV2 is its packaging capacity, being limited to about 4kb, which prevent packaging of a full-length *CFTR* cDNA with a strong promoter.

7.2.6 HOST IMMUNE RESPONSE

Recent studies from 2019 showed presence of Cas9 existing antibodies. Because of *Staphylococcus. aureus* presence in humans, the study showed that 78% of people had pre-existing SaCas9. Concerning *Streptococcus pyogenes*, the study showed that 58% of people had pre-existing antibodies against SpCas9. Those antibodies could be a problem for Cas9 insertion in the body, creating an immune response that might be harmful for the patient, and/or could decrease the efficiency of the treatment (228).

7.2.7 PRIME EDITING

In December 2019, a new gene editing technique called prime editing has been described by the group of David Liu (229). This technique uses a nickase Cas9 such as the one used for base editing, fused to an engineered reverse transcriptase. The Cas9 forms a complex with a prime editing gRNA (pegRNA) that contains the modification of interest. The nickase breaks one DNA strand, making a 3'-hydroxyl group available as a primer for reverse transcription, using the 3'-end of the pegRNA as a template (Figure 7.1). Using prime editing, the paper showed an average of editing efficiency from 20% to 50% in a range of cell types, with 1 to 10% indel formation (229). A current limitation of this technique consists on the size of the mutation to be corrected. Indeed, because the correction in located in the pegRNA, only mutations that are up to 34bp in length can currently be corrected. However, those mutations correspond to about 89% of the variants in the ClinVar database.

It was not possible to use prime editing for this thesis to try and correct W1282X, but this technique was still evaluated for its feasibility in the laboratory, in HEK cells, with a control pegRNA, which should create a precise 3bp 5'-CTT-3' insertion in the *HEK3* gene (see appendix, figure S3). The preliminary results showed efficient 5'-CTT-3' insertion, with very low, if any, indels formation.

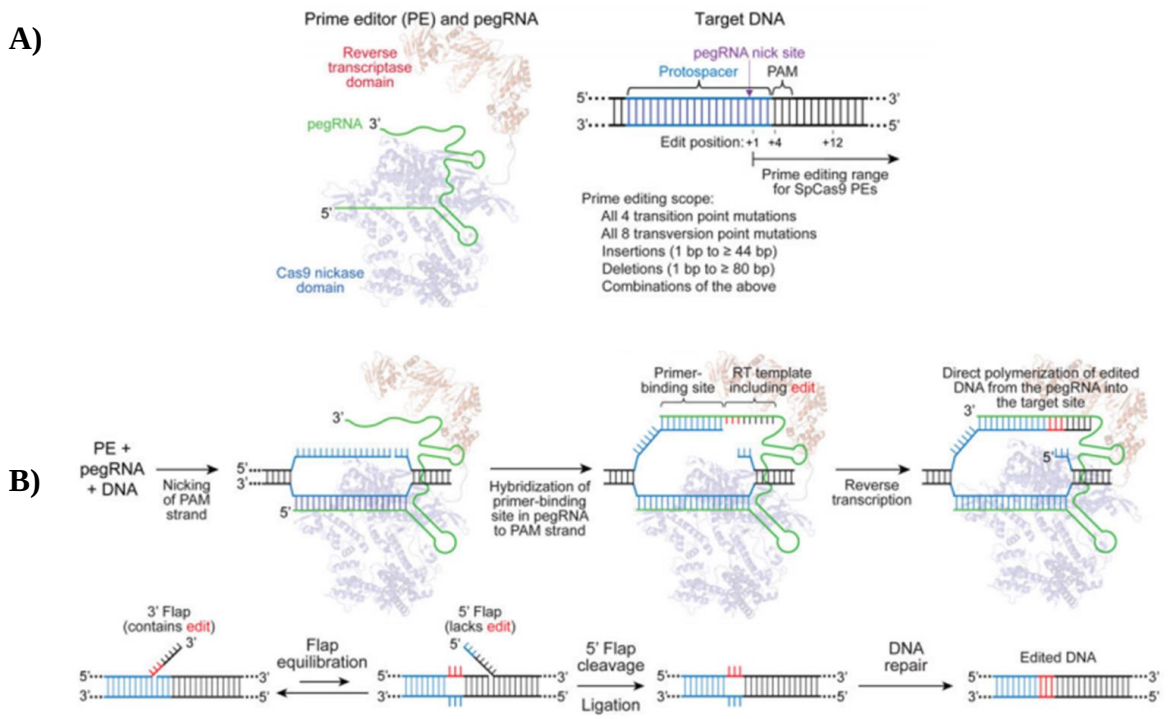


Figure 7.1: Principle of the prime editing technique. A) Structure of the prime editing complex. The RNP complex consists of a prime editing gRNA (pegRNA) bound to a nickase Cas9 containing a reverse transcriptase domain fused to it. B) The RNP complex recognises the target DNA and creates a nick into the non-target strand. The DNA part upstream the cut site hybridises with the pegRNA inducing polymerisation of the DNA, using the pegRNA-containing modifications as template. The 5' DNA flap is cleaved and after ligation and DNA repair, the DNA possesses the wanted modifications.
***(Figure adapted from Anzalone et al. 2019;** (229)(229).*

Figure 7.1 - Principle of the prime editing technique

Importantly, finding a gene editing technique that could correct CFTR in airways would hopefully help patients to have a less cumbersome treatment. As stated above, if airway cells can be successfully corrected in the lungs, this correction could last 1 to few years, until the next cell turnover. This could make patients' life easier, alleviating the treatments that they have to take, and possibly making the treatments cheaper. A restoration of CFTR production is expected to lead to a fluidification of the mucus, as well as a decrease in lung infection. However, depending on the mutation, some organs like the pancreas could be already damaged from birth, and even CFTR correction will not be able to correct this defect. Therefore, patients will still have to take many pills in order to have a normal nutrition. Even though CFTR correction in not a cure for CF, it is hoped that it could help making patients' life much easier and increase their survival.

REFERENCES

1. Cystic Fibrosis Trust. What is cystic fibrosis? 2017 [Available from: https://www.cysticfibrosis.org.uk/what-is-cystic-fibrosis (accessed 02 June 2020).
2. Genetics Home Reference. Cystic fibrosis 2012 [Available from: https://ghr.nlm.nih.gov/condition/cystic-fibrosis#resources (accessed 02 June 2020).
3. Bobadilla JL, Macek M, Jr., Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. Hum Mutat. 2002;19(6):575-606.
4. Farrell PM. The prevalence of cystic fibrosis in the European Union. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society. 2008;7(5):450-3.
5. ANDERSEN DH. CYSTIC FIBROSIS OF THE PANCREAS AND ITS RELATION TO CELIAC DISEASE: A CLINICAL AND PATHOLOGIC STUDY. American Journal of Diseases of Children. 1938;56(2):344-99.
6. Elborn JS. Personalised medicine for cystic fibrosis: treating the basic defect. Eur Respir Rev. 2013;22(127):3-5.
7. Mantle DJ, Norman AP. Life-table for cystic fibrosis. British medical journal. 1966;2(5524):1238-41.
8. Fogarty A, Hubbard R, Britton J. International comparison of median age at death from cystic fibrosis. Chest. 2000;117(6):1656-60.
9. Cystic Fibrosis Foundation. Understanding Changes in Life Expectancy 2020 [Available from: https://www.cff.org/Research/Researcher-Resources/Patient-Registry/Understanding-Changes-in-Life-Expectancy/ (accessed 02 June 2020).
10. Penketh AR, Wise A, Mearns MB, Hodson ME, Batten JC. Cystic fibrosis in adolescents and adults. Thorax. 1987;42(7):526-32.
11. Stoltz DA, Meyerholz DK, Welsh MJ. Origins of cystic fibrosis lung disease. The New England journal of medicine. 2015;372(4):351-62.
12. Farber S, Shwachman H, Maddock CL. PANCREATIC FUNCTION AND DISEASE IN EARLY LIFE. I. PANCREATIC ENZYME ACTIVITY AND THE CELIAC SYNDROME. The Journal of clinical investigation. 1943;22(6):827-38.
13. Kopelman H, Corey M, Gaskin K, Durie P, Weizman Z, Forstner G. Impaired chloride secretion, as well as bicarbonate secretion, underlies the fluid secretory defect in the cystic fibrosis pancreas. Gastroenterology. 1988;95(2):349-55.
14. Chryssostalis A, Hubert D, Coste J, Kanaan R, Burgel PR, Desmazes-Dufeu N, et al. Liver disease in adult patients with cystic fibrosis: a frequent and independent prognostic factor associated with death or lung transplantation. Journal of hepatology. 2011;55(6):1377-82.
15. Lanng S, Thorsteinsson B, Nerup J, Koch C. Influence of the development of diabetes mellitus on clinical status in patients with cystic fibrosis. European journal of pediatrics. 1992;151(9):684-7.
16. Denning CR, Sommers SC, Quigley HJ, Jr. Infertility in male patients with cystic fibrosis. Pediatrics. 1968;41(1):7-17.
17. Darling RC, Disant'Agnese PA, Perera GA, Andersen DH. Electrolyte abnormalities of the sweat in fibrocystic disease of the pancreas. The American journal of the medical sciences. 1953;225(1):67-70.
18. Di Sant'Agnese PA, Darling RC, Perera GA, Shea E. Abnormal electrolyte composition of sweat in cystic fibrosis of the pancreas; clinical significance and relationship to the disease. Pediatrics. 1953;12(5):549-63.
19. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet. 2015;16(1):45-56.
20. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science (New York, NY). 1989;245(4922):1073-80.
21. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science (New York, NY). 1989;245(4922):1066-73.

22. Hefferon TW, Broackes-Carter FC, Harris A, Cutting GR. Atypical 5' splice sites cause CFTR exon 9 to be vulnerable to skipping. Am J Hum Genet. 2002;71(2):294-303.
23. Cheng SH, Gregory RJ, Marshall J, Paul S, Souza DW, White GA, et al. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. Cell. 1990;63(4):827-34.
24. Gregory RJ, Cheng SH, Rich DP, Marshall J, Paul S, Hehir K, et al. Expression and characterization of the cystic fibrosis transmembrane conductance regulator. Nature. 1990;347(6291):382-6.
25. Farinha CM, Swiatecka-Urban A, Brautigan DL, Jordan P. Regulatory Crosstalk by Protein Kinases on CFTR Trafficking and Activity. Frontiers in chemistry. 2016;4:1.
26. O'Riordan CR, Lachapelle AL, Marshall J, Higgins EA, Cheng SH. Characterization of the oligosaccharide structures associated with the cystic fibrosis transmembrane conductance regulator. Glycobiology. 2000;10(11):1225-33.
27. Sheppard DN, Welsh MJ. Structure and function of the CFTR chloride channel. Physiological reviews. 1999;79(1 Suppl):S23-45.
28. Hwang TC, Kirk KL. The CFTR ion channel: gating, regulation, and anion permeation. Cold Spring Harbor perspectives in medicine. 2013;3(1):a009498.
29. Knowles M, Gatzy J, Boucher R. Increased bioelectric potential difference across respiratory epithelia in cystic fibrosis. The New England journal of medicine. 1981;305(25):1489-95.
30. Frizzell RA, Hanrahan JW. Physiology of epithelial chloride and fluid secretion. Cold Spring Harbor perspectives in medicine. 2012;2(6):a009563.
31. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket SE, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature. 2018;560(7718):319-24.
32. Plasschaert LW, Žilionis R, Choo-Wing R, Savova V, Knehr J, Roma G, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature. 2018;560(7718):377-81.
33. Tarran R. Regulation of airway surface liquid volume and mucus transport by active ion transport. Proceedings of the American Thoracic Society. 2004;1(1):42-6.
34. Borowitz D. CFTR, bicarbonate, and the pathophysiology of cystic fibrosis. Pediatric pulmonology. 2015;50 Suppl 40:S24-s30.
35. Garland AL, Walton WG, Coakley RD, Tan CD, Gilmore RC, Hobbs CA, et al. Molecular basis for pH-dependent mucosal dehydration in cystic fibrosis airways. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(40):15973-8.
36. Tarran R, Redinbo MR. Mammalian short palate lung and nasal epithelial clone 1 (SPLUNC1) in pH-dependent airway hydration. The international journal of biochemistry & cell biology. 2014;52:130-5.
37. Saint-Criq V, Gray MA. Role of CFTR in epithelial physiology. Cellular and molecular life sciences : CMLS. 2017;74(1):93-115.
38. Hill DB, Long RF, Kissner WJ, Atieh E, Garbarine IC, Markovetz MR, et al. Pathological mucus and impaired mucus clearance in cystic fibrosis patients result from increased concentration, not altered pH. The European respiratory journal. 2018;52(6).
39. Stutts MJ, Knowles MR, Gatzy JT, Boucher RC. Oxygen consumption and ouabain binding sites in cystic fibrosis nasal epithelium. Pediatr Res. 1986;20(12):1316-20.
40. Bear CE, Li CH, Kartner N, Bridges RJ, Jensen TJ, Ramjeesingh M, et al. Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR). Cell. 1992;68(4):809-18.
41. Tsui LC, Dorfman R. The cystic fibrosis gene: a molecular genetic perspective. Cold Spring Harbor perspectives in medicine. 2013;3(2):a009472.
42. Cystic Fibrosis Centre. Cystic Fibrosis Mutation Database 2011 [Available from: http://www.genet.sickkids.on.ca/cftr/StatisticsPage.html (accessed 02 June 2020).
43. Human Genome Variation Society. Nomenclature for the description of sequence variants 2016 [Available from: http://www.hgvs.org/mutnomen/ (accessed 02 June 2020).
44. Castellani C. CFTR2: How will it help care? Paediatric respiratory reviews. 2013;14 Suppl 1:2-5.
45. Boyle MP, De Boeck K. A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect. The Lancet Respiratory medicine. 2013;1(2):158-63.
46. De Boeck K, Amaral MD. Classification of CFTR mutation classes - Authors' reply. The Lancet Respiratory medicine. 2016;4(8):e39.

47. Kerem E. Pharmacological induction of CFTR function in patients with cystic fibrosis: mutation-specific therapy. Pediatric pulmonology. 2005;40(3):183-96.
48. Cystic Fibrosis Trust. What are the causes of cystic fibrosis? 2019 [Available from: https://www.cysticfibrosis.org.uk/what-is-cystic-fibrosis/what-causes-cystic-fibrosis (accessed 02 June 2020).
49. De Boeck K, Amaral MD. Progress in therapies for cystic fibrosis. The Lancet Respiratory medicine. 2016;4(8):662-74.
50. Zainal Abidin N, Haq IJ, Gardner AI, Brodlie M. Ataluren in cystic fibrosis: development, clinical studies and where are we now? Expert opinion on pharmacotherapy. 2017;18(13):1363-71.
51. Hyde SC, Southern KW, Gileadi U, Fitzjohn EM, Mofford KA, Waddell BE, et al. Repeat administration of DNA/liposomes to the nasal epithelium of patients with cystic fibrosis. Gene therapy. 2000;7(13):1156-65.
52. Moss RB, Milla C, Colombo J, Accurso F, Zeitlin PL, Clancy JP, et al. Repeated aerosolized AAV-CFTR for treatment of cystic fibrosis: a randomized placebo-controlled phase 2B trial. Human gene therapy. 2007;18(8):726-32.
53. Yonemitsu Y, Kitson C, Ferrari S, Farley R, Griesenbach U, Judd D, et al. Efficient gene transfer to airway epithelium using recombinant Sendai virus. Nature biotechnology. 2000;18(9):970-3.
54. Zabner J, Couture LA, Gregory RJ, Graham SM, Smith AE, Welsh MJ. Adenovirus-mediated gene transfer transiently corrects the chloride transport defect in nasal epithelia of patients with cystic fibrosis. Cell. 1993;75(2):207-16.
55. Vallières E, Elborn JS. Cystic fibrosis gene mutations: evaluation and assessment of disease severity. Advances in Genomics and Genetics. 2014;4:161-72.
56. Cystic Fibrosis Foundation. Drug Development pipeline 2018 [Available from: https://www.cff.org/trials/pipeline (accessed 02 June 2020).
57. Cystic Fibrosis Foundation. CFTR Modulator Therapies 2019 [Available from: https://www.cff.org/Life-With-CF/Treatments-and-Therapies/Medications/CFTR-Modulator-Therapies/ (accessed 10 September 2020).
58. Vertex Pharmaceuticals Incorporated. Highlights of Prescribing Information 2019 [Available from: https://pi.vrtx.com/files/uspi_ivacaftor.pdf (accessed 10 September 2020).
59. Vertex Pharmaceuticals Incorporated. WHO SYMDEKO® IS FOR 2020 [Available from: https://www.symdeko.com/who-symdeko-is-for (accessed 10 September 2020).
60. Vertex Pharmaceuticals Incorporated. kalydeco 2020 [Available from: https://www.kalydeco.com/ (accessed 02 June 2020).
61. Andrew Joseph. New Vertex cystic fibrosis drug approved, extending treatments to 90% of patients: STAT; 2019 [Available from: https://www.statnews.com/2019/10/21/new-vertex-cystic-fibrosis-drug-approved-extending-treatments-to-90-of-patients/ (accessed 02 June 2020).
62. Withers N. Pricing dispute over cystic fibrosis drug Orkambi continues in the UK2019 02 June 2020. Available from: https://www.europeanpharmaceuticalreview.com/news/83645/pricing-dispute-orkambi/ (accessed 02 June 2020).
63. Angelis A, Kanavos P, López-Bastida J, Linertová R, Nicod E, Serrano-Aguilar P. Social and economic costs and health-related quality of life in non-institutionalised patients with cystic fibrosis in the United Kingdom. BMC Health Serv Res. 2015;15:428.
64. Abou Alaiwa MH, Launspach JL, Grogan B, Carter S, Zabner J, Stoltz DA, et al. Ivacaftor-induced sweat chloride reductions correlate with increases in airway surface liquid pH in cystic fibrosis. JCI insight. 2018;3(15).
65. American Society of Gene & Cell Therapy. Gene Therapy Basics 2020 [Available from: https://www.asgct.org/education/gene-therapy-basics (accessed 02 June 2020).
66. Griesenbach U, Alton EW. Progress in gene and cell therapy for cystic fibrosis lung disease. Curr Pharm Des. 2012;18(5):642-62.
67. Griesenbach U, Pytel KM, Alton EW. Cystic Fibrosis Gene Therapy in the UK and Elsewhere. Human gene therapy. 2015;26(5):266-75.
68. Loring HS, ElMallah MK, Flotte TR. Development of rAAV2-CFTR: History of the First rAAV Vector Product to be Used in Humans. Hum Gene Ther Methods. 2016;27(2):49-58.
69. Alton E, Armstrong DK, Ashby D, Bayfield KJ, Bilton D, Bloomfield EV, et al. Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial. The Lancet Respiratory medicine. 2015;3(9):684-91.

70. Crystal RG, McElvaney NG, Rosenfeld MA, Chu CS, Mastrangeli A, Hay JG, et al. Administration of an adenovirus containing the human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis. Nature genetics. 1994;8(1):42-51.
71. European Medicines Agency. Imlygic 2019 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/imlygic (accessed 10 September 2020).
72. European Medicines Agency. Strimvelis 2020 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/strimvelis (accessed 10 September 2020).
73. European Medicines Agency. Kymriah 2020 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/kymriah (accessed 10 September 2020).
74. European Medicines Agency. Yescarta 2020 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/yescarta (accessed 10 September 2020).
75. European Medicines Agency. Luxturna 2020 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/luxturna (accessed 10 September 2020).
76. European Medicines Agency. Glybera 2017 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/glybera (accessed 10 September 2020).
77. Warner E. Goodbye Glybera! The World's First Gene Therapy will be Withdrawn: Labiotech.eu; 2017 [Available from: https://www.labiotech.eu/medical/uniqure-glybera-marketing-withdrawn/ (accessed 02 June 2020).
78. Aiuti A, Cossu G, de Felipe P, Galli MC, Narayanan G, Renner M, et al. The committee for advanced therapies' of the European Medicines Agency reflection paper on management of clinical risks deriving from insertional mutagenesis. Hum Gene Ther Clin Dev. 2013;24(2):47-54.
79. Kim YG, Chandrasegaran S. Chimeric restriction endonuclease. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(3):883-7.
80. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(3):1156-60.
81. Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res. 2000;28(17):3361-9.
82. Genetics Home Reference. What are genome editing and CRISPR-Cas9? 2020 [Available from: https://ghr.nlm.nih.gov/primer/genomicresearch/genomeediting (accessed 04 June 2020).
83. Kim H, Kim JS. A guide to genome engineering with programmable nucleases. Nat Rev Genet. 2014;15(5):321-34.
84. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol. 1987;169(12):5429-33.
85. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151(Pt 8):2551-61.
86. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science (New York, NY). 2007;315(5819):1709-12.
87. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012;482(7385):331-8.
88. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics. 2007;8:172.
89. Wei Y, Chesne MT, Terns RM, Terns MP. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. Nucleic Acids Res. 2015;43(3):1749-58.
90. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nature reviews Microbiology. 2020;18(2):67-83.
91. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (New York, NY). 2012;337(6096):816-21.
92. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science (New York, NY). 2014;343(6176):1247997.

93. Nowak CM, Lawson S, Zerez M, Bleris L. Guide RNA engineering for versatile Cas9 functionality. Nucleic Acids Res. 2016;44(20):9555-64.
94. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;507(7490):62-7.
95. Shou J, Li J, Liu Y, Wu Q. Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. Mol Cell. 2018;71(4):498-509.e4.
96. Lemos BR, Kaplan AC, Bae JE, Ferrazzoli AE, Kuo J, Anand RP, et al. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(9):E2040-e7.
97. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology. 2013;31(9):822-6.
98. Zheng T, Hou Y, Zhang P, Zhang Z, Xu Y, Zhang L, et al. Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. Sci Rep. 2017;7:40638.
99. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature biotechnology. 2014;32(3):279-84.
100. Wang X, Wang Y, Wu X, Wang J, Wang Y, Qiu Z, et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nature biotechnology. 2015;33(2):175-8.
101. Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, et al. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 2013;41(19):9049-61.
102. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature. 2017;551(7681):464-71.
103. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015;523(7561):481-5.
104. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science (New York, NY). 2016;351(6268):84-8.
105. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529(7587):490-5.
106. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature. 2017;550(7676):407-10.
107. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nature biotechnology. 2018;36(3):265-71.
108. Manghwar H, Li B, Ding X, Hussain A, Lindsey K, Zhang X, et al. CRISPR/Cas Systems in Genome Editing: Methodologies and Tools for sgRNA Design, Off-Target Evaluation, and Strategies to Mitigate Off-Target Effects. Adv Sci (Weinh). 2020;7(6):1902312.
109. Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y, et al. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell. 2016;165(4):949-62.
110. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163(3):759-71.
111. Swarts DC, Jinek M. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. Mol Cell. 2019;73(3):589-600.e4.
112. Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nature biotechnology. 2016;34(8):863-8.
113. Yamano T, Zetsche B, Ishitani R, Zhang F, Nishimasu H, Nureki O. Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. Mol Cell. 2017;67(4):633-45.e3.
114. Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nature biotechnology. 2016;34(8):869-74.
115. Wang Y, Liu KI, Sutrisnoh NB, Srinivasan H, Zhang J, Li J, et al. Systematic evaluation of CRISPR-Cas systems reveals design principles for genome editing in human cells. Genome biology. 2018;19(1):62.

116. Tóth E, Weinhardt N, Bencsura P, Huszár K, Kulcsár PI, Tálas A, et al. Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. Biol Direct. 2016;11:46.
117. Gao L, Cox DBT, Yan WX, Manteiga JC, Schneider MW, Yamano T, et al. Engineered Cpf1 variants with altered PAM specificities. Nature biotechnology. 2017;35(8):789-92.
118. Integrated DNA Technologies. Alt-R A.s. Cas12a (Cpf1) *Ultra* nuclease: Integrated DNA Technologies,; 2019 [Available from: https://sfvideo.blob.core.windows.net/sitefinity/docs/default-source/flyer/alt-r-a-s-cas12a-(cpf1)-ultra-nuclease1974506494.pdf?sfvrsn=80ca1807_2 (accessed 28 September 2020).
119. Shahbazi R, Sghia-Hughes G, Reid JL, Kubek S, Haworth KG, Humbert O, et al. Targeted homology-directed repair in blood stem and progenitor cells with CRISPR nanoformulations. Nature materials. 2019;18(10):1124-32.
120. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst). 2008;7(10):1765-71.
121. Ranjha L, Howard SM, Cejka P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. Chromosoma. 2018;127(2):187-214.
122. Anand R, Ranjha L, Cannavo E, Cejka P. Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. Mol Cell. 2016;64(5):940-50.
123. Fell VL, Schild-Poulter C. The Ku heterodimer: function in DNA repair and beyond. Mutat Res Rev Mutat Res. 2015;763:15-29.
124. Mimori T, Hardin JA. Mechanism of interaction between Ku protein and DNA. The Journal of biological chemistry. 1986;261(22):10375-9.
125. Blier PR, Griffith AJ, Craft J, Hardin JA. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. The Journal of biological chemistry. 1993;268(10):7594-601.
126. Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol. 2003;4(9):712-20.
127. Chang HH, Watanabe G, Gerodimos CA, Ochi T, Blundell TL, Jackson SP, et al. Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. The Journal of biological chemistry. 2016;291(47):24377-89.
128. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017;18(8):495-506.
129. Goodarzi AA, Yu Y, Riballo E, Douglas P, Walker SA, Ye R, et al. DNA-PK autophosphorylation facilitates Artemis endonuclease activity. Embo j. 2006;25(16):3880-9.
130. Gu J, Li S, Zhang X, Wang LC, Niewolik D, Schwarz K, et al. DNA-PKcs regulates a single-stranded DNA endonuclease activity of Artemis. DNA Repair (Amst). 2010;9(4):429-37.
131. Kowalczykowski SC. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. Cold Spring Harb Perspect Biol. 2015;7(11).
132. Zhu Z, Chung WH, Shim EY, Lee SE, Ira G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. Cell. 2008;134(6):981-94.
133. Tran PT, Erdeniz N, Dudley S, Liskay RM. Characterization of nuclease-dependent functions of Exo1p in Saccharomyces cerevisiae. DNA Repair (Amst). 2002;1(11):895-912.
134. Bae SH, Choi E, Lee KH, Park JS, Lee SH, Seo YS. Dna2 of Saccharomyces cerevisiae possesses a single-stranded DNA-specific endonuclease activity that is able to act on double-stranded DNA in the presence of ATP. The Journal of biological chemistry. 1998;273(41):26880-90.
135. Pinto C, Kasaciunaite K, Seidel R, Cejka P. Human DNA2 possesses a cryptic DNA unwinding activity that functionally integrates with BLM or WRN helicases. Elife. 2016;5.
136. Jensen RB, Carreira A, Kowalczykowski SC. Purified human BRCA2 stimulates RAD51-mediated recombination. Nature. 2010;467(7316):678-83.
137. Dion V, Kalck V, Horigome C, Towbin BD, Gasser SM. Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. Nat Cell Biol. 2012;14(5):502-9.
138. Miné-Hattab J, Rothstein R. Increased chromosome mobility facilitates homology search during recombination. Nat Cell Biol. 2012;14(5):510-7.
139. Forget AL, Kowalczykowski SC. Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. Nature. 2012;482(7385):423-7.

140. Qi Z, Redding S, Lee JY, Gibb B, Kwon Y, Niu H, et al. DNA sequence alignment by microhomology sampling during homologous recombination. Cell. 2015;160(5):856-69.
141. Anand R, Beach A, Li K, Haber J. Rad51-mediated double-strand break repair and mismatch correction of divergent substrates. Nature. 2017;544(7650):377-80.
142. Eggler AL, Inman RB, Cox MM. The Rad51-dependent pairing of long DNA substrates is stabilized by replication protein A. The Journal of biological chemistry. 2002;277(42):39280-8.
143. Mason JM, Dusad K, Wright WD, Grubb J, Budke B, Heyer WD, et al. RAD54 family translocases counter genotoxic effects of RAD51 in human tumor cells. Nucleic Acids Res. 2015;43(6):3180-96.
144. Barber LJ, Youds JL, Ward JD, McIlwraith MJ, O'Neil NJ, Petalcorin MI, et al. RTEL1 maintains genomic stability by suppressing homologous recombination. Cell. 2008;135(2):261-71.
145. Bugreev DV, Brosh RM, Jr., Mazin AV. RECQ1 possesses DNA branch migration activity. The Journal of biological chemistry. 2008;283(29):20231-42.
146. Mitchel K, Lehner K, Jinks-Robertson S. Heteroduplex DNA position defines the roles of the Sgs1, Srs2, and Mph1 helicases in promoting distinct recombination outcomes. PLoS Genet. 2013;9(3):e1003340.
147. Nimonkar AV, Sica RA, Kowalczykowski SC. Rad52 promotes second-end DNA capture in double-stranded break repair to form complement-stabilized joint molecules. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(9):3077-82.
148. Duckett DR, Murchie AI, Diekmann S, von Kitzing E, Kemper B, Lilley DM. The structure of the Holliday junction, and its resolution. Cell. 1988;55(1):79-89.
149. Swuec P, Costa A. Molecular mechanism of double Holliday junction dissolution. Cell Biosci. 2014;4:36.
150. Wyatt HD, West SC. Holliday junction resolvases. Cold Spring Harb Perspect Biol. 2014;6(9):a023192.
151. Bizard AH, Hickson ID. The dissolution of double Holliday junctions. Cold Spring Harb Perspect Biol. 2014;6(7):a016477.
152. Vaidyanathan S, Salahudeen AA, Sellers ZM, Bravo DT, Choi SS, Batish A, et al. High-Efficiency, Selection-free Gene Repair in Airway Stem Cells from Cystic Fibrosis Patients Rescues CFTR Function in Differentiated Epithelia. Cell stem cell. 2020;26(2):161-71.e4.
153. Berical A, Lee RE, Randell SH, Hawkins F. Challenges Facing Airway Epithelial Cell-Based Therapy for Cystic Fibrosis. Front Pharmacol. 2019;10:74.
154. Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature. 2016;540(7631):144-9.
155. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533(7603):420-4.
156. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature. 2017;551(7681):464-71.
157. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nature biotechnology. 2018;36(8):765-71.
158. Lee C, Hyun Jo D, Hwang GH, Yu J, Kim JH, Park SE, et al. CRISPR-Pass: Gene Rescue of Nonsense Mutations Using Adenine Base Editors. Molecular therapy : the journal of the American Society of Gene Therapy. 2019;27(8):1364-71.
159. Mention K, Santos L, Harrison PT. Gene and Base Editing as a Therapeutic Option for Cystic Fibrosis-Learning from Other Diseases. Genes. 2019;10(5).
160. Yasui M, Suenaga E, Koyama N, Masutani C, Hanaoka F, Gruz P, et al. Miscoding properties of 2'-deoxyinosine, a nitric oxide-derived DNA Adduct, during translesion synthesis catalyzed by human DNA polymerases. J Mol Biol. 2008;377(4):1015-23.
161. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nature biotechnology. 2018;36(9):843-6.
162. Kunz C, Saito Y, Schär P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. Cellular and molecular life sciences : CMLS. 2009;66(6):1021-38.
163. Standage-Beier K, Zhang Q, Wang X. Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. ACS Synth Biol. 2015;4(11):1217-25.

164. Slotkin W, Nishikura K. Adenosine-to-inosine RNA editing and human disease. Genome Med. 2013;5(11):105.
165. Ryu SM, Koo T, Kim K, Lim K, Baek G, Kim ST, et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nature biotechnology. 2018;36(6):536-9.
166. Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science (New York, NY). 2018;361(6408):1259-62.
167. Huang S, Liao Z, Li X, Liu Z, Li G, Li J, et al. Developing ABEmax-NG with Precise Targeting and Expanded Editing Scope to Model Pathogenic Splice Site Mutations In Vivo. iScience. 2019;15:640-8.
168. Grünewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature. 2019;569(7756):433-7.
169. Grünewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nature biotechnology. 2019;37(9):1041-8.
170. Li H, Haurigot V, Doyon Y, Li T, Wong SY, Bhagwat AS, et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature. 2011;475(7355):217-21.
171. Bednarski C, Tomczak K, Vom Hövel B, Weber WM, Cathomen T. Targeted Integration of a Super-Exon into the CFTR Locus Leads to Functional Correction of a Cystic Fibrosis Cell Line Model. PloS one. 2016;11(8):e0161072.
172. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. The Journal of general virology. 1977;36(1):59-74.
173. Domingue JC, Ao M, Sarathy J, George A, Alrefai WA, Nelson DJ, et al. HEK-293 cells expressing the cystic fibrosis transmembrane conductance regulator (CFTR): a model for studying regulation of Cl- transport. Physiological reports. 2014;2(9).
174. Cozens AL, Yezzi MJ, Kunzelmann K, Ohrui T, Chin L, Eng K, et al. CFTR expression and chloride secretion in polarized immortal human bronchial epithelial cells. American journal of respiratory cell and molecular biology. 1994;10(1):38-47.
175. Valley HC, Bukis KM, Bell A, Cheng Y, Wong E, Jordan NJ, et al. Isogenic cell models of cystic fibrosis-causing variants in natively expressing pulmonary epithelial cells. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society. 2019;18(4):476-83.
176. Aksit MA, Bowling AD, Evans TA, Joynt AT, Osorio D, Patel S, et al. Decreased mRNA and protein stability of W1282X limits response to modulator therapy. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society. 2019;18(5):606-13.
177. van Kuppeveld FJ, van der Logt JT, Angulo AF, van Zoest MJ, Quint WG, Niesters HG, et al. Genus- and species-specific identification of mycoplasmas by 16S rRNA amplification. Applied and environmental microbiology. 1993;59(2):655.
178. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 2019;47(W1):W171-w4.
179. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281-308.
180. BioLabs. NE. NEBioCalculator™ v1.10.1. 2020, June 02.
181. Deskgen. TIDE: Tracking of Indels by DEcomposition. 2020, June 02.
182. Synthego. ICE Analysis. 2020, June 02.
183. Kluesner MG, Nedveck DA, Lahr WS, Garbe JR, Abrahante JE, Webber BR, et al. EditR: A Method to Quantify Base Editing from Sanger Sequencing. Crispr j. 2018;1(3):239-50.
184. Harrison PT, Sanz DJ, Hollywood JA. Impact of gene editing on the study of cystic fibrosis. Human genetics. 2016;135(9):983-92.
185. Schubert M, Thommandru, B. & Wang, B. Want to achieve higher rates of homology-directed repair in your CRISPR HDR experiment? 2018 [Available from: https://www.idtdna.com/pages/education/decoded/article/want-to-achieve-higher-rates-of-homology-directed-repair-(hdr)-in-your-crispr-experiment (accessed 08 September 2020).
186. Lin H, Li H, Cho HJ, Bian S, Roh HJ, Lee MK, et al. Air-liquid interface (ALI) culture of human bronchial epithelial cell monolayers as an in vitro model for airway drug transport studies. J Pharm Sci. 2007;96(2):341-50.

187. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science (New York, NY). 2016;353(6305).
188. Huang S, Liao Z, Li X, Liu Z, Li G, Li J, et al. Developing ABEmax-NG with Precise Targeting and Expanded Editing Scope to Model Pathogenic Splice Site Mutations In Vivo. iScience. 2019;15:640-8.
189. Hamosh A, Rosenstein BJ, Cutting GR. CFTR nonsense mutations G542X and W1282X associated with severe reduction of CFTR mRNA in nasal epithelial cells. Human molecular genetics. 1992;1(7):542-4.
190. Geurts MH, de Poel E, Amatngalim GD, Oka R, Meijers FM, Kruisselbrink E, et al. CRISPR-Based Adenine Editors Correct Nonsense Mutations in a Cystic Fibrosis Organoid Biobank. Cell stem cell. 2020;26(4):503-10.e7.
191. Grunewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature. 2019;569(7756):433-7.
192. Grunewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nature biotechnology. 2019;37(9):1041-8.
193. Sanz DJ, Hollywood JA, Scallan MF, Harrison PT. Cas9/gRNA targeted excision of cystic fibrosis-causing deep-intronic splicing mutations restores normal splicing of CFTR mRNA. PloS one. 2017;12(9):e0184009.
194. Chang J, Wang T, Nian R, Lau A, Hoi KM, Ho SC, et al. Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. mAbs. 2015;7(2):403-12.
195. ViralZone. Ribosomal skipping 2008 [Available from: https://viralzone.expasy.org/914 (accessed 03 June 2020).
196. Michelini F, Pitchiaya S, Vitelli V, Sharma S, Gioia U, Pessina F, et al. Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. Nat Cell Biol. 2017;19(12):1400-11.
197. Holstein M, Mesa-Nuñez C, Miskey C, Almarza E, Poletti V, Schmeer M, et al. Efficient Non-viral Gene Delivery into Human Hematopoietic Stem Cells by Minicircle Sleeping Beauty Transposon Vectors. Molecular therapy : the journal of the American Society of Gene Therapy. 2018;26(4):1137-53.
198. Hornstein BD, Roman D, Arévalo-Soliz LM, Engevik MA, Zechiedrich L. Effects of Circular DNA Length on Transfection Efficiency by Electroporation into HeLa Cells. PloS one. 2016;11(12):e0167537.
199. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Research. 2005;33(suppl_2):W526-W31.
200. Bednarski C, Tomczak K, Vom Hovel B, Weber WM, Cathomen T. Targeted Integration of a Super-Exon into the CFTR Locus Leads to Functional Correction of a Cystic Fibrosis Cell Line Model. PloS one. 2016;11(8):e0161072.
201. Gillen AE, Gosalia N, Leir SH, Harris A. MicroRNA regulation of expression of the cystic fibrosis transmembrane conductance regulator gene. Biochem J. 2011;438(1):25-32.
202. Megiorni F, Cialfi S, Dominici C, Quattrucci S, Pizzuti A. Synergistic post-transcriptional regulation of the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) by miR-101 and miR-494 specific binding. PloS one. 2011;6(10):e26601.
203. Oglesby IK, Chotirmall SH, McElvaney NG, Greene CM. Regulation of cystic fibrosis transmembrane conductance regulator by microRNA-145, -223, and -494 is altered in ΔF508 cystic fibrosis airway epithelium. J Immunol. 2013;190(7):3354-62.
204. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215-33.
205. Byrne SM, Ortiz L, Mali P, Aach J, Church GM. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic Acids Res. 2015;43(3):e21.
206. Burge C. Splicing of precursors to mRNAs by the spliceosomes. Cold Spring Harbor Laboratory Press. 1999.
207. Aix Marseille Universite. Human Splicing Finder 2013 [Available from: http://www.umd.be/HSF/ (accessed 31 May 2020).

208. Suzuki S, Crane AM, Anirudhan V, Barillà C, Matthias N, Randell SH, et al. Highly Efficient Gene Editing of Cystic Fibrosis Patient-Derived Airway Basal Cells Results in Functional CFTR Correction. Molecular therapy : the journal of the American Society of Gene Therapy. 2020;28(7):1684-95.
209. Chen S, Sun S, Moonen D, Lee C, Lee AY, Schaffer DV, et al. CRISPR-READI: Efficient Generation of Knockin Mice by CRISPR RNP Electroporation and AAV Donor Infection. Cell Rep. 2019;27(13):3780-9.e4.
210. Jiang T, Henderson JM, Coote K, Cheng Y, Valley HC, Zhang XO, et al. Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. Nature communications. 2020;11(1):1979.
211. Zafra MP, Schatoff EM, Katti A, Foronda M, Breinig M, Schweitzer AY, et al. Optimized base editors enable efficient editing in cells, organoids and mice. Nature biotechnology. 2018;36(9):888-93.
212. Glasgow AMA, De Santi C, Greene CM. Non-coding RNA in cystic fibrosis. Biochem Soc Trans. 2018;46(3):619-30.
213. Johnson LG, Boyles SE, Wilson J, Boucher RC. Normalization of raised sodium absorption and raised calcium-mediated chloride secretion by adenovirus-mediated expression of cystic fibrosis transmembrane conductance regulator in primary human cystic fibrosis airway epithelial cells. The Journal of clinical investigation. 1995;95(3):1377-82.
214. Ramalho AS, Beck S, Meyer M, Penque D, Cutting GR, Amaral MD. Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. American journal of respiratory cell and molecular biology. 2002;27(5):619-27.
215. Chou SJ, Yang P, Ban Q, Yang YP, Wang ML, Chien CS, et al. Dual Supramolecular Nanoparticle Vectors Enable CRISPR/Cas9-Mediated Knockin of Retinoschisin 1 Gene-A Potential Nonviral Therapeutic Solution for X-Linked Juvenile Retinoschisis. Adv Sci (Weinh). 2020;7(10):1903432.
216. Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods. 2011;8(9):753-5.
217. Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M. Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol. 1998;18(1):93-101.
218. Yao X, Zhang M, Wang X, Ying W, Hu X, Dai P, et al. Tild-CRISPR Allows for Efficient and Precise Gene Knockin in Mouse and Human Cells. Dev Cell. 2018;45(4):526-36.e5.
219. Rawlins EL, Hogan BL. Ciliated epithelial cell lifespan in the mouse trachea and lung. American journal of physiology Lung cellular and molecular physiology. 2008;295(1):L231-4.
220. Gaudelli NM, Lam DK, Rees HA, Solá-Esteves NM, Barrera LA, Born DA, et al. Directed evolution of adenine base editors with increased activity and therapeutic application. Nature biotechnology. 2020.
221. Stoltz DA, Rokhlina T, Ernst SE, Pezzulo AA, Ostedgaard LS, Karp PH, et al. Intestinal CFTR expression alleviates meconium ileus in cystic fibrosis pigs. The Journal of clinical investigation. 2013;123(6):2685-93.
222. Sun X, Yi Y, Yan Z, Rosen BH, Liang B, Winter MC, et al. In utero and postnatal VX-770 administration rescues multiorgan disease in a ferret model of cystic fibrosis. Sci Transl Med. 2019;11(485).
223. Alton EW, Beekman JM, Boyd AC, Brand J, Carlon MS, Connolly MM, et al. Preparation for a first-in-man lentivirus trial in patients with cystic fibrosis. Thorax. 2017;72(2):137-47.
224. Flotte TR, Fischer AC, Goetzmann J, Mueller C, Cebotaru L, Yan Z, et al. Dual reporter comparative indexing of rAAV pseudotyped vectors in chimpanzee airway. Molecular therapy : the journal of the American Society of Gene Therapy. 2010;18(3):594-600.
225. Gao G, Vandenberghe LH, Alvira MR, Lu Y, Calcedo R, Zhou X, et al. Clades of Adeno-associated viruses are widely disseminated in human tissues. J Virol. 2004;78(12):6381-8.
226. Naso MF, Tomkowicz B, Perry WL, 3rd, Strohl WR. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. BioDrugs. 2017;31(4):317-34.
227. Manno CS, Pierce GF, Arruda VR, Glader B, Ragni M, Rasko JJ, et al. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nature medicine. 2006;12(3):342-7.

228. Charlesworth CT, Deshpande PS, Dever DP, Camarena J, Lemgart VT, Cromer MK, et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. Nature medicine. 2019;25(2):249-54.
229. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149-57.